SENSING DICTIONARY CONSTRUCTION FOR ORTHOGONAL MATCHING PURSUIT ALGORITHM IN COMPRESSIVE SENSING

SENSING DICTIONARY CONSTRUCTION FOR ORTHOGONAL MATCHING PURSUIT ALGORITHM IN COMPRESSIVE SENSING

BY

BO LI, M.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2013) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE:	SENSING DICTIONARY CONSTRUCTION FOR ORTHOGONAL MATCHING PURSUIT ALGORITHM IN COMPRESSIVE SENSING
AUTHOR:	Bo Li M.Sc. (Electronic and Communication Engineering) Harbin Institute of Technology, Harbin, China
SUPERVISOR:	Dr. T.Kirubarajan
NUMBER OF PAGES:	xii, 62

To my supervisor, my parents and my grandmother

Abstract

In compressive sensing, the fundamental problem is to reconstruct sparse signal from its nonadaptive insufficient linear measurement. Besides sparse signal reconstruction algorithms, measurement matrix or measurement dictionary plays an important part in sparse signal recovery. Orthogonal Matching Pursuit (OMP) algorithm, which is widely used in compressive sensing, is especially affected by measurement dictionary. Measurement dictionary with small restricted isometry constant or coherence could improve the performance of OMP algorithm. Based on measurement dictionary, sensing dictionary can be constructed and can be incorporated into OMP algorithm. In this thesis, two methods are proposed to design sensing dictionary. In the first method, sensing dictionary design problem is formulated as a linear programming problem. The solution is unique and can be obtained by standard linear programming method such as primal-dual interior point method. The major drawback of linear programming based method is its high computational complexity. The second method is termed sensing dictionary designing algorithm. In this algorithm, each atom of sensing dictionary is designed independently to reduce the maximal magnitude of its inner product with measurement dictionary. Compared with linear programming based method, the proposed sensing dictionary design algorithm is of low computational complexity and the performance is similar. Simulation results

indicate that both of linear programming based method and the proposed sensing dictionary designing algorithm can design sensing dictionary with small mutual coherence and cumulative coherence. When the designed sensing dictionary is applied to OMP algorithm, the performance of OMP algorithm improves.

Acknowledgements

First and foremost I would like to pay sincere thanks to my supervisor, Dr. Kiruba, for supporting me to finish the master degree for the past two years. Dr. Kiruba really impresses me not only by his academic achievement but also by his nice and glorious personality.

I appreciate and would like to thank Dr. T. Field and Dr. Patriciu for being my committee member. Sincere thanks to Dr. Reilly, Dr. Bakr and Dr. Bruce for the patient discussion with me.

I am grateful to Dr. Shirani and Dr. Jeremic for their kindly suggestions and corrections for my thesis. Thanks for the energy and patient time spent on my thesis.

Special thanks is given to Christina Dalima, who modified my thesis so carefully that I was really moved.

Last but not least, I would also like to thank Mrs. Cheryl Gies, who gave me numberless help patiently and friendly from the first day I came to McMaster.

Notation

$\mathbf{x}$	sparse signal
$\mathbf{\Phi}$	measurement dictionary
φ_i	ith column of measurement dictionary
$\mathbf{y}$	measurement signal
$\delta_{\mathbf{k}}$	Isometry Restricted Constant
μ	coherence
$\mu_1(k)$	kth order cumulative coherence
μ_w	Welch bound
$\mathbf{r}$	residual signal
Λ	index set of selected atoms in OMP algorithm
Γ	index set of nonzero components of sparse signal
$\mathbf{P}$	orthogonal projection operator
$\mathbf{P}^{\perp}$	orthogonal projection complementary operator
$\mathbf{h}$	identification vector
Ψ	sensing dictionary
ψ_i	ith column of sensing dictionary
δ'_k	generalized Restricted Isometry Constant

μ'	mutual coherence
$\mu'_1(k)$	kth order cumulative mutual coherence
$\mathbf{G}$	type Gram matrix
$\mathbf{H}$	ideal type Gram matrix
$\mathtt{G}$	set of type Gram matrix
$\mathtt{H}$	set of ideal type Gram matrix
$\mathbf{W}$	weighting matrix
$\mathbf{d}_i$	directional vector
$\mathbf{d'}_i$	directional vector
α	step length

Contents

List of Figures

Chapter 1

Introduction and Problem Statement

1.1 Introduction to Compressive Sensing

Compressive sensing is a new signal processing technique, which attracts the attention of researchers in various areas such as radar, communication and image processing. The basic idea behind compressive sensing is that an unknown sparse signal can be accurately recovered from under determined nonadaptive linear measurement. Here, sparse signal means that the number of nonzero components is far less than the length of signal itself. Under determined nonadaptive linear measurement represents linear measurement with the number of measurement being less than the length of sparse signal. Also, the linear measurement is independent of the sparse signal. In general, compressive sensing model could be written in the following form:

$$\mathbf{y} = \mathbf{\Phi x} \tag{1.1}$$

where $\mathbf{x} \in \mathbf{R}^{n\times 1}$ is sparse signal with the number of nonzero components k far less than its length n. $\mathbf{\Phi} \in \mathbf{R}^{m\times n}$ is measurement matrix with its number of rows m being less than that of columns n. Measurement matrix is often called measurement dictionary with each column called atom. $\mathbf{y} \in \mathbf{R}^{m\times 1}$ is measurement signal.

Generally speaking, equation (1.1) has infinite number of solutions as it is an under determined linear equation. One simple solution is Moore-Penrose pseudoinverse solution with its closed form as $\mathbf{\Phi}\dagger\,\mathbf{y}$, which is based on the least square criterion. $\mathbf{\Phi}\dagger$ is Moor-Penrose pseudoinverse of $\mathbf{\Phi}$. However, least square criterion often produces dense signal rather than sparse signal, although the simple solution of closed form exists.

If we make the assumption that the signal $\mathbf{x}$ is sparse, intuitively we could choose the one which is sparsest among the infinite number of solutions as recovered signal. This naturally leads to the following optimization problem:

$$\min_{\mathbf{x}\in\mathbf{R}^{n\times 1}} \|\mathbf{x}\|_0 \quad s.t \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x} \tag{1.2}$$

where $\|\cdot\|_0$ is pseudo zero norm which counts the number of nonzero components of the input vector. Problem (1.2) is difficult to solve as it is an NP hard problem. More generally, solving problem (1.2) involves total number of combination $\binom{n}{m}$, where m is the length of measurement signal. Greedy algorithm is an important class of algorithms which aims to solve the l_0 minimization problem (1.2) approximately. Classical representatives are *Matching Pursuit* algorithm (MP), *Orthogonal Matching Pursuit* algorithm (OMP) and *Subspace Pursuit* algorithm(SP).

Moreover, l_0 problem could be relaxed to l_p $(0 < p < 1)$ or even l_1 problem under

some conditions. l_p and l_1 problems can be written as

$$\min_{\mathbf{x}\in\mathbf{R}^{n\times 1}} \|\mathbf{x}\|_p \quad s.t \quad \mathbf{y} = \mathbf{\Phi x} \tag{1.3}$$

and

$$\min_{\mathbf{x}\in\mathbf{R}^{n\times 1}} \|\mathbf{x}\|_1 \quad s.t \quad \mathbf{y} = \mathbf{\Phi x} \tag{1.4}$$

where $\|\cdot\|_p$ and $\|\cdot\|_1$ are p norm and one norm of the input vector respectively.

l_p problem is not a convex problem. However, some algorithms are employed to solve it such as FOCUSS (Gorodnitsky and Rao, 1997) and MCCR (Mourad and Reilly, 2010). l_1 problem is easier to solve than l_p as the former one is a convex problem, which means that there exists an unique global solution (Candes and Tao, 2005). Classical algorithm solving the l_1 problem is *Basis Pursuit* algorithm(BP).

The equivalence of l_0 problem with l_p or l_1 problem requires restrictions of measurement dictionary and sparsity of original sparse signal. Here, sparsity is defined as the number of nonzero components of sparse signal. One popular parameter describing the property of measurement dictionary is *Restricted Isomentry Constant* (RIC) (Candes and Tao, 2005), which is defined as following:

Defination 1.1.1. *(Candes and Tao, 2005): For measurement dictionary, if the following inequality holds true for any sparse signal* $\mathbf{x}$ *with sparsity* k

$$\sqrt{1-\delta_k}\,\|\mathbf{x}\|_2 \leq \|\mathbf{\Phi x}\|_2 \leq \sqrt{1+\delta_k}\,\|\mathbf{x}\|_2 \tag{1.5}$$

the measurement dictionary is said to be with k*th order Restricted Isomentry Constant(RIC)*

δ_k.

For l_1 minimization problem (1.4), if $2k$th order RIC of measurement dictionary satisfies $\delta_{2k} \leq \sqrt{2} - 1$, the solution of l_0 minimization problem (1.2) and that of l_1 problem (1.4) are identical (Candes and Tao, 2005).

A more fundamental problem is whether the solution of l_0 problem (1.2) is equal to the original signal. Using the language of RIC, it can be said that when $\delta_{2k} < 1$ the solution to (1.2) is exactly the original signal.

Another important parameter is called *coherence* and *culmulate coherence* of measurement dictionary (Tropp, 2004), which are defined as following:

Defination 1.1.2. *(Tropp, 2004) The coherence and kth order cumulative coherence of measurement dictionary* $\mathbf{\Phi}$ *are defined as*

$$\mu = \max_{1 \leq i,j \leq n, i \neq j} |\langle \varphi_i, \varphi_j \rangle| \tag{1.6}$$

$$\mu_1(k) = \max_{|\Gamma|=k, 1 \leq i,j \leq n} \max_{i \notin \Gamma, j \in \Gamma} \sum |\langle \varphi_i, \varphi_j \rangle| \tag{1.7}$$

where $\varphi_i \in \mathbf{R}^{m \times 1}$ *represents the ith* $(1 \leq i \leq n)$ *column of measurement dictionary,* Γ *is the set of indices and* $|\cdot|$ *gives cardinality of input set.*

It is proved that if $\mu < \frac{1}{2(k-1)}$ or $\mu_1(k-1) + \mu_1(k) < 1$, solution of l_0 problem (1.2) and that of l_1 problem (1.4) are identical. Moreover, both of Baisis Pursuit algorithm and OMP algorithm are able to recover the sparse signal in compressive sensing (Tropp, 2004).

Generally, there are various kinds of sparse signal recovery algorithms such as the ones based on greedy atom selection (Tropp and Gilbert, 2007), convex optimization (Chen *et al.*, 1998), Bayesian inference (Ji *et al.*, 2008) and mixture of

them (Huggins and Zucker, 2007). We do not intend to introduce them here as it is a rather big topic and beyond the scope of this thesis.

Compressive sensing has many applications in signal processing field and beyond. It is successfully applied to direction of arrival(DOA) estimation (Malioutov *et al.*, 2005), single pixel imaging (Duarte *et al.*, 2008), imaging and video processing (Donoho, 2006). In DOA estimation problem, the measurement dictionary is composed by dense samples from the array manifold with each angle predefined. The number of targets is assumed to be sparse compared to the number of angles of the samples(or the number of measurement dictionary). Measurement signal can be viewed as a linear combination of samples in the measurement dictionary with noise added. As such, DOA estimation problem is transferred to compressive sensing problem.

1.2 Problem Formulation

From section 1.1, it can be seen that RIC or coherence represents the quality of measurement dictionary. Generally speaking, measurement dictionary with small RIC is favored in compressive sensing (Candes and Tao, 2005). When RIC is equal to zero exactly, linear measurement using measurement dictionary actually satisfies Parseval theorem. However, as measurement dictionary is a redundant dictionary, which means that the number of column is larger than that of rows, RIC could never become zero.

Following the same logic, measurement dictionary with small coherence or cumulative coherence is favored in compressive sensing (Tropp, 2004). In the extreme case when both of coherence and cumulative coherence become zero, measurement

dictionary becomes an orthogonal matrix and sparse signal recovery becomes simple. However, we could never expect that coherence becomes zero as measurement dictionary is redundant. It is proved that for redundant measurement dictionary, coherence coefficient can not be reduced to be zero. There exists a lower bound of coherence μ_w which is called Welch bound. For real dictionary with size m by n and $n \leq 0.5m\,(m-1)$, Welch bound is

$$\mu_w = \sqrt{\frac{n-m}{m\,(n-1)}} \tag{1.8}$$

Note that Welch bound is a theoretical bound. The bound can not be always achieved by any pair of m and n. Welch bound tells us the best we could do to reduce the coherence of measurement dictionary. (Tropp *et al.*, 2005) gives a method to design measurement dictionary with small coherence based on alternating projection method. From the simulation in (Tropp *et al.*, 2005), it can be seen that for dictionary with certain dimensions, the coherence could achieve Welch bound.

As mentioned in section 1.1, RIC is often used to evaluate the quality of measurement dictionary. However, for a given measurement dictionary, calculation of its kth order RIC is impractical as it involves computation of eigenvalue of Gram matrix of the matrix formed by any k columns of measurement dictionary. kth order RIC satisfies the following equation:

$$\delta_k = \max_{|\Gamma|=k} \left\{ \lambda_{max}\left(\mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma\right) - 1, 1 - \lambda_{min}\left(\mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma\right) \right\} \tag{1.9}$$

where $\mathbf{\Phi}_\Gamma$ is a matrix composed of atoms of $\mathbf{\Phi}$ with index in the set Γ, $\lambda_{max}\left(\mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma\right)$ and $\lambda_{min}\left(\mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma\right)$ are maximal and minimal eigenvalue of $\mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma$.

Moreover, the relationship between RIC and cumulative coherence can be written as

$$\delta_k \leq \mu_1 (k-1) \tag{1.10}$$

A simple interpretation of (1.10) is that kth order RIC is equal to or less than $k-1$th order cumulative coherence. Although calculation of RIC is computational intensive, calculation of cumulative coherence is relatively easy. Therefore, most of dictionary construction algorithms aim to construct dictionaries with small coherence or cumulative coherence. RIC is often used to analyze the performance of signal recovery algorithm (Davenport.M.A and Wakin.M.B, 2010). We could see from (1.10) that when $k-1$th order cumulative coherence is small, kth order RIC will never get very large. Thus, it is reasonable to use coherence as a criterion to construct measurement dictionary.

OMP algorithm, as introduced in detail in the next chapter, is a classical greedy algorithm that is widely used in compressive sensing. OMP algorithm is easy to understand and the computational complexity is not large. OMP algorithm sequentially selects atom to calculate the sparse signal. There are mainly three steps in one loop of OMP algorithm: atom identification, coefficient renewal and residual update. The most important step in OMP algorithm is atom identification. In atom identification step, atom is selected based on its correlation with residual signal. As will be analyzed in the next chapter, coherence or cumulative coherence has great influence on the performance of atom identification. When measurement dictionary is orthogonal and the measurement is noise free, true atoms can be always selected based on the maximum correlation criterion. In practice, measurement dictionary is redundant,

which is far from orthogonal matrix. In this case, it is proved that measurement dictionary with small coherence will improve the ability of correct atom selection in atom identification step. Therefore, measurement dictionary with small coherence has a positive effect on the performance of OMP algorithm.

In common practice, atom identification in OMP algorithm is based on correlation of residual signal and measurement dictionary. Firstly, inner product of residual signal and measurement dictionary is calculated. Index corresponding to the maximum absolute value of the inner product is viewed as index of selected atom. This procedure can be expressed as

$$i = \max_{1 \leq j \leq n} \left| \varphi_j^T \mathbf{r} \right| \tag{1.11}$$

where i is the index of selected atom and $\mathbf{r} \in \mathbf{R}^{m \times 1}$ is residual signal. In (Schnass and Vandergheynst, 2008), a new concept termed sensing dictionary is proposed which is used to correlate with residual signal in atom identification step. With the use of sensing dictionary, the identification step can be written as

$$i = \max_{1 \leq j \leq n} \left| \psi_j^T \mathbf{r} \right| \tag{1.12}$$

where $\psi_j \in \mathbf{R}^{m \times 1}$ is jth atom of sensing dictionary $\mathbf{\Psi} \in \mathbf{R}^{m \times n}$ and others are the same as (1.11).

The advantage of sensing dictionary involved here is that the mutual coherence of sensing and measurement dictionary is less than coherence of measurement dictionary. The definition of mutual coherence is similar to that of coherence and the exact definition will be given in the next chapter. Similarly, it is proved that small mutual

coherence will improve the ability of OMP algorithm to select a correct atom in atom identification step.

In (Schnass and Vandergheynst, 2008), sensing dictionary is designed via alternating projection method. The basic idea is to construct a sensing dictionary to make the Gram type matrix approximate the set of predefined matrix. The set of predefined matrix is a set which contains matrix with diagonal entries equal to one and the absolute value of non diagonal elements equal to or less than Welch bound. Alternating projection method is utilized to minimize the approximation. As such, sensing dictionary could be designed and the corresponding mutual coherence will be reduced.

In (Li *et al.*, 2011), a pair of sensing and measurement dictionary is constructed simultaneously based on alternating projection method. The method used in (Li *et al.*, 2011) is similar to that in (Schnass and Vandergheynst, 2008). The difference is that in (Li *et al.*, 2011) both sensing and measurement dictionaries are updated iteratively and mutual coherence could be lowered further. Sensing dictionary could be applied to any kind of measurement dictionary, from typical Gaussian dictionary to the pairs of orthogonal dictionaries. In contrast, simultaneous construction of sensing and measurement dictionaries gives no choice of the form of measurement dictionary and its application is limited.

1.3 Contribution of the Thesis

In this thesis, we give two methods to design sensing dictionary based on given measurement dictionary.

In the first algorithm, the sensing dictionary design problem is formulated as an

optimization problem with linear equality and inequalities constraints. This optimization problem is then formulated as standard linear programming problem which can be solved by classical linear programming methods such as log barrier interior point method and primal-dual interior point method. In this thesis, primal-dual interior point algorithm is utilized, which is in line with the Matlab function `linprog.m`. When the size of sensing dictionary is large, primal-dual interior algorithm is computationally demanding.

The second algorithm, which is termed sensing dictionary design algorithm, is introduced to reduce the computational complexity while keeping good performance. To construct each atom of sensing dictionary, the proposed sensing dictionary design algorithm is to minimize the maximal magnitude of its inner product with measurement dictionary iteratively.

The two algorithms given in this thesis are with similar performance while the latter is computationally less intensive. Both of the algorithms could reduce the value of mutual coherence and cumulative mutual coherence of sensing and measurement dictionary. When applying the designed sensing dictionary to OMP algorithm, the sparse signal successful recovery rate in compressive sensing is improved.

1.4 Outline of the Thesis

In chapter one, a basic introduction of compressive sensing problem and related concepts are given. In chapter two, OMP algorithm is analyzed using RIC. More over, the concept of sensing dictionary in OMP algorithm is introduced and the advantage of sensing dictionary involved in OMP algorithm is given. Chapter three gives two kinds of algorithm to construct sensing dictionary. In both of the algorithms, each

column of sensing dictionary is designed separately. In the first algorithm, designing each atom of sensing dictionary can be formulated as an infinite norm optimization problem with linear equation and inequations constraints. This problem is further written as linear programming problem which can be solved by standard linear programming method. The second method is to minimize the largest coherence of each column of sensing dictionary with measurement dictionary iteratively. The minimization is implemented iteratively and mutual coherence coefficient is minimized consequently. Chapter four gives simulation results using the methods proposed in chapter three. The performance of OMP algorithm using different kind of sensing dictionaries is also compared. Chapter five gives the summary of the thesis.

Chapter 2

Analysis of Orthogonal Matching Pursuit Algorithm

2.1 Introduction to Orthogonal Matching Pursuit Algorithm

For sparse signal $\mathbf{x}$ with sparsity k, denote Γ as set of index corresponding to the nonzero components of sparse signal. Sparsity k also means $|\Gamma| = k$. There is

$$\mathbf{x}(i) \neq 0 \quad i \in \Gamma \tag{2.1}$$

where $\mathbf{x}(i)$ is the ith component of sparse signal x.

The basic compressive sensing problem can be written as

$$\mathbf{y} = \mathbf{\Phi x} = \sum_{i \in \Gamma} \varphi_i \mathbf{x}\,(i) \tag{2.2}$$

where $\varphi_i \in \mathbf{R}^{m\times 1}$ is ith column or atom of measurement dictionary. As can be seen from (2.2) that measurement signal $\mathbf{y}$ is a linear combination of k atoms of measurement dictionary. Moreover, the atoms and combination coefficients are determined by nonzero components of sparse signal $\mathbf{x}$.

Orthogonal Matching Pursuit(OMP) algorithm is an iterative greedy algorithm that solves the problem (2.2). OMP algorithm is given in the table 2.1 where $\mathbf{\Phi}_\Lambda$ represents the matrix formed by columns of $\mathbf{\Phi}$ with index in the set Λ. At each iteration, OMP algorithm selects the atom of measurement dictionary $\mathbf{\Phi}$ which is most correlated with the current residual $\mathbf{r}$ as Step 1 in table 2.1. In Step 2, the index of this atom is added into the set of selected atoms Λ. The algorithm updates the nonzero coefficient of sparse signal using least square technique. In Step 3, the residual signal is renewed using the coefficient and index set estimated in Step 2. Compared with other iterative methods, a major advantage of the OMP algorithm is its simplicity and fast implementation.

In OMP algorithm, once one atom is selected and its index is added to the index set Λ, residual signal is updated. Let $\mathbf{P}_\Lambda$ and $\mathbf{P}_\Lambda^\perp$ be the orthogonal projection operator on the column space of $\mathbf{\Phi}_\Lambda$ and its orthogonal complement respectively. $\mathbf{P}_\Lambda$ and $\mathbf{P}_\Lambda^\perp$ can be written explicitly as

$$\mathbf{P}_\Lambda = \mathbf{\Phi}_\Lambda \mathbf{\Phi}_\Lambda^\dagger \tag{2.3}$$

$$\mathbf{P}_\Lambda^\perp = \mathbf{I} - \mathbf{P}_\Lambda = \mathbf{I} - \mathbf{\Phi}_\Lambda \mathbf{\Phi}_\Lambda^\dagger \tag{2.4}$$

where $\mathbf{\Phi}_\Lambda^\dagger = \left(\mathbf{\Phi}_\Lambda^T \mathbf{\Phi}_\Lambda\right)^{-1} \mathbf{\Phi}_\Lambda^T$ is the Moore-Penrose pseudoinverse of $\mathbf{\Phi}_\Lambda$. As such, the

Input: measurement dictionary $\mathbf{\Phi}$, measurement signal $\mathbf{y}$ and sparsity k.

Initialization: index set $\Lambda=\emptyset$ and residual signal $\mathbf{r} = \mathbf{y}$.

repeat step 1 to 3 for k iterations.

Step1(identification): $i = \max_{1\leq j\leq n} \left|\varphi_j^T \mathbf{r}\right|$.

Step2(sparse signal reconstruction): $\Lambda = \Lambda \cup i$, $\tilde{\mathbf{x}} = \mathbf{\Phi}_\Lambda^\dagger \mathbf{y}$.

Step3(residual update): $\mathbf{r} = \mathbf{r} - \Phi_\Lambda \tilde{\mathbf{x}}$.

Output: reconstructed sparse signal $\hat{\mathbf{x}}$ with $\hat{\mathbf{x}}_\Lambda = \tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}_{\Lambda^c} = \mathbf{0}$.

Table 2.1: Orthogonal Matching Pursuit

residual signal can be written as

$$\mathbf{r} = \mathbf{y} - \mathbf{\Phi}_\Lambda \tilde{\mathbf{x}} = \mathbf{y} - \mathbf{\Phi}_\Lambda \mathbf{\Phi}_\Lambda^\dagger \mathbf{y} = \left(\mathbf{I} - \mathbf{P}_\Lambda\right) \mathbf{y} = \mathbf{P}_\Lambda^\perp \mathbf{y} \tag{2.5}$$

As can be seen from expression (2.5), residual signal is the orthogonal complementary projection on the subspace spanned by all the selected atoms whose index is in the set Λ. According to the property of orthogonal and orthogonal complementary projection, residual signal $\mathbf{r}$ is orthogonal to the selected atoms which can be expressed as

$$\varphi_i^T \mathbf{r} = 0 \qquad i \in \Lambda \tag{2.6}$$

In the atom identification step, the identification vector $\mathbf{h}$ can be written as

$$\mathbf{h} = \mathbf{\Phi}^T \mathbf{r} \tag{2.7}$$

(2.6) can be written in another way as

$$\mathbf{h}_\Lambda = \mathbf{\Phi}_\mathbf{\Lambda}^\mathbf{T} \mathbf{r} = \mathbf{0}_{|\Lambda| \times 1} \tag{2.8}$$

where $\mathbf{h}_\Lambda$ denotes the components of vector $\mathbf{h}$ whose index is in the index set Λ and $\mathbf{0}_{|\Lambda| \times 0}$ is a zero vector with size $|\Lambda|$ by 1. As can be seen from (2.8), once one atom is selected and its index is added to the index set Λ, its inner product with the residual signal is always zero. It also means that the selected atoms could not be selected twice.

Another important fact is that in OMP algorithm, only when all the right atoms are selected, the reconstructed sparse signal is exact in the noiseless case. This could be further proved and analyzed in the next section.

2.2 Analysis of OMP Algorithm

The performance of OMP algorithm is heavily dependent on measurement dictionary. Usually the performance of OMP algorithm is analyzed using *RIC* or coherence. In the first iteration of OMP algorithm, the residual signal can be written explicitly as

$$\mathbf{r} = \mathbf{\Phi x} \tag{2.9}$$

According to the *RIC* of measurement dictionary, the following inequalities can

be obtained (Davenport.M.A and Wakin.M.B, 2010):

$$|\mathbf{h}(i) - \mathbf{x}(i)| \leq \delta_{k+1} \|\mathbf{x}\|_2 \quad i \in \Gamma \tag{2.10}$$

$$|\mathbf{h}(i)| \leq \delta_{k+1} \|\mathbf{x}\|_2 \quad i \notin \Gamma \tag{2.11}$$

where $\mathbf{h}(i)$ is ith component of identification vector $\mathbf{h}$. (2.10) and (2.11) mean that the identification vector $\mathbf{h}$ is analogous to the original signal $\mathbf{x}$ but experiences changes whose magnitude is bounded. If the maximum value of $\mathbf{h}(j), j \in \Gamma$ is smaller than the maximum value of $\mathbf{h}(j), j \notin \Gamma$, OMP algorithm will definitely fail in the first iteration. Thus, we expect that the changes are small and OMP algorithm could select a right atom.

After some iterations, assume that the selected atoms are all right, which means that $\Lambda \subset \Gamma$. If we want to represent residual signal as a linear combination of atoms of measurement dictionary, it could be written as

$$\begin{aligned}
\mathbf{r} &= (\mathbf{I} - \mathbf{P}_\Lambda)\mathbf{y} = (\mathbf{I} - \mathbf{P}_\Lambda)\mathbf{\Phi}\mathbf{x} = (\mathbf{I} - \mathbf{P}_\Lambda)\left(\mathbf{\Phi}_\Lambda \mathbf{x}_\Lambda + \mathbf{\Phi}_{\Gamma\setminus\Lambda}\mathbf{x}_{\Gamma\setminus\Lambda}\right) \\
&= (\mathbf{I} - \mathbf{P}_\Lambda)\mathbf{\Phi}_{\Gamma\setminus\Lambda}\mathbf{x}_{\Gamma\setminus\Lambda} = \mathbf{\Phi}_{\Gamma\setminus\Lambda}\mathbf{x}_{\Gamma\setminus\Lambda} - \mathbf{\Phi}_\Lambda\left(\mathbf{\Phi}_\Lambda^T\mathbf{\Phi}_\Lambda\right)^{-1}\mathbf{\Phi}_\Lambda^T\mathbf{\Phi}_{\Gamma\setminus\Lambda}\mathbf{x}_{\Gamma\setminus\Lambda} \\
&= \begin{pmatrix} \mathbf{\Phi}_{\Gamma\setminus\Lambda} & \mathbf{\Phi}_\Lambda \end{pmatrix} \begin{pmatrix} \mathbf{x}_{\Gamma\setminus\Lambda} \\ -\left(\mathbf{\Phi}_\Lambda^T\mathbf{\Phi}_\Lambda\right)^{-1}\mathbf{\Phi}_\Lambda^T\mathbf{\Phi}_{\Gamma\setminus\Lambda}\mathbf{x}_{\Gamma\setminus\Lambda} \end{pmatrix}
\end{aligned} \tag{2.12}$$

where the third equality means that measurement signal could be decomposed as two parts: one part is a linear combination of atoms whose index is in the set Λ and the other part is a linear combination of atoms whose index is in the set $\Gamma \setminus \Lambda$.

From (2.12), we could see that when the selected atoms in set Λ is right, the estimated nonzero coefficient, which is $\tilde{\mathbf{x}} = \mathbf{\Phi}_\Lambda^\dagger \mathbf{y}$, is not always right. The reason is

that if the estimated coefficient $\tilde{\mathbf{x}}$ are right which is equal to $\mathbf{x}_\Lambda$, the residual signal $\mathbf{r}$ is only a linear combination of the unselected right atoms, which can be written as

$$\mathbf{r} = \mathbf{y} - \mathbf{\Phi}_\Lambda \tilde{\mathbf{x}} = \mathbf{\Phi}\mathbf{x} - \mathbf{\Phi}_\Lambda \mathbf{x}_\Lambda = \mathbf{\Phi}_{\Gamma\backslash\Lambda}\mathbf{x}_{\Gamma\backslash\Lambda} \tag{2.13}$$

However, as can be seen from (2.12), residual signal $\mathbf{r}$ is not only the linear combination of the unselected atoms. Residual signal $\mathbf{r}$ is a linear combination of all the right atoms whose index is in set Γ.

In the ideal situation, we hope that when right atoms are selected, its corresponding coefficients can be estimated correctly. As such, the residual signal is simply the linear combination of the unselected right atoms such as

$$\mathbf{r} = \mathbf{\Phi}_{\Gamma\backslash\Lambda}\mathbf{x}_{\Gamma\backslash\Lambda} \tag{2.14}$$

However, (2.14) is seldom true. From (2.12), if we want to write residual signal $\mathbf{r}$ as a linear combination of atoms whose index is in set Γ, which is $\mathbf{r} = \mathbf{\Phi}_\Gamma \mathbf{x}_r$, linear combination coefficient $\mathbf{x}_r \in \mathbf{R}^{|\Gamma|\times 1}$ can be written as

$$\mathbf{x}_r = \begin{pmatrix} \mathbf{x}_{\Gamma\backslash\Lambda} \\ -\left(\mathbf{\Phi}_\Lambda^T \mathbf{\Phi}_\Lambda\right)^{-1} \mathbf{\Phi}_\Lambda^T \mathbf{\Phi}_{\Gamma\backslash\Lambda}\mathbf{x}_{\Gamma\backslash\Lambda} \end{pmatrix} \tag{2.15}$$

Comparing with ideal case (2.14), besides $\mathbf{x}_{\Gamma\backslash\Lambda}$, there is another part as

$$-\left(\mathbf{\Phi}_\Lambda^T \mathbf{\Phi}_\Lambda\right)^{-1} \mathbf{\Phi}_\Lambda^T \mathbf{\Phi}_{\Gamma\backslash\Lambda}\mathbf{x}_{\Gamma\backslash\Lambda}$$

which is not always zero. Lemma 2.1 gives the relationship between $\mathbf{x}_r$ and $\mathbf{x}_{\Gamma\backslash\Lambda}$.

Lemma 2.1. *For* $\mathbf{x}_r$ *defined in (2.12), the following inequality exists:*

$$\|\mathbf{x}_r\|_2 \leq \frac{1}{1-\delta_{k+1}} \left\|\mathbf{x}_{\Gamma\backslash\Lambda}\right\|_2 \tag{2.16}$$

Lemma 2.1 tells us that in the signal $\mathbf{x}_r$, although the coefficients corresponding to the selected atom are not zero, two norm of $\mathbf{x}_r$ is bounded by that of the unidentified coefficients. Moreover, when *RIC* δ_{k+1} is small enough, the coefficient $\mathbf{x}_r$ approximate the coefficient of unselected atoms $\mathbf{x}_{\Gamma\backslash\Lambda}$. In other word, two norm of the part $-\left(\mathbf{\Phi}_\Lambda^T\mathbf{\Phi}_\Lambda\right)^{-1}\mathbf{\Phi}_\Lambda^T\mathbf{\Phi}_{\Gamma\backslash\Lambda}\mathbf{x}_{\Gamma\backslash\Lambda}$ in $\mathbf{x}_r$ is small.

In OMP algorithm, after certain iterations, the relationship between the identity vector $\mathbf{h}$ and the signal $\mathbf{x}_{\Gamma\backslash\Lambda}$ can be easily obtained as:

$$|\mathbf{h}(i) - \mathbf{x}(i)| \leq \frac{\delta_{k+1}}{1-\delta_{k+1}} \left\|\mathbf{x}_{\Gamma\backslash\Lambda}\right\|_2 \qquad i \in \Gamma \setminus \Lambda \tag{2.17}$$

$$|\mathbf{h}(i)| = 0 \qquad i \in \Lambda \tag{2.18}$$

$$|\mathbf{h}(i)| \leq \frac{\delta_{k+1}}{1-\delta_{k+1}} \left\|\mathbf{x}_{\Gamma\backslash\Lambda}\right\|_2 \qquad i \in \Gamma^c \tag{2.19}$$

where $\Gamma^c = \{1, 2, \cdots, n\} \setminus \Gamma$. (2.17) to (2.19) are similar to (2.10) and (2.11) respectively and we do not give additional discussion here. (2.18) actually corresponds to (2.8), which means that the elements of identification vector corresponding to index of the selected atom are always zero. This is also one of the most important characteristics of OMP algorithm that one atom could never be selected twice.

In compressive sensing, there are two types of sparse signals that are frequently used. For the first type, its nonzero component is Gaussian distributed or with other distribution. The nonzero component of the second type is of unit value. It is widely recognized that the first type signal is easier to recover than the second type for

greedy algorithm. (2.17),(2.18) and (2.19) can give the explanation of this phenomena. For both nonzero and zero components of the sparse signal, the upper bound of perturbations are the same which is proportional to the square root of the energy of signal $\left\|\mathbf{x}_{\Gamma\backslash\Lambda}\right\|_2$. Assume that the energy of signal $\mathbf{x}_{\Gamma\backslash\Lambda}$ is fixed. In other word, the maximum magnitude of perturbation is fixed. For nonzero component of sparse signal, the larger the maximal magnitude, the higher possibility that its corresponding component of identification vector being the maximal over all the components. In this case, OMP algorithm will choose the corresponding atom. Conversely, when the magnitude of all the nonzero component of sparse signal with equal value, none of the magnitude of nonzero component is significant and the largest magnitude of $\mathbf{h}$ would not be very high. When OMP algorithm selects the atom index according to the largest magnitude of $\mathbf{h}$, the selection is not right with high probability if RIC δ_{k+1} is not small enough. Thus, recovering sparse signal with its nonzero entries of equal magnitude is regarded as challenge for not only OMP algorithm but also for all the greedy algorithms. The analysis here keeps in line with observations.

Assume OMP algorithm selects the right atom at each iteration. After k iterations, all the atoms whose index set is Γ are selected and the estimated coefficients are

$$\hat{\mathbf{x}} = \left(\mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma\right)^{-1} \mathbf{\Phi}_\Gamma^T \mathbf{y} = \left(\mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma\right)^{-1} \mathbf{\Phi}_\Gamma^T \mathbf{\Phi}_\Gamma \mathbf{x} = \mathbf{x} \tag{2.20}$$

as can be seen from (2.20) that when all the right atoms are selected in the noiseless case, the spare signal can be recovered successfully.

Another question is what the value of RIC should be to guarantee the success of OMP algorithm. The following theorem will give the answer to this question.

Theorem 2.1. *(Mo and Shen, 2012) Suppose that RIC δ_{k+1} of measurement dictionary $\mathbf{\Phi}$ satisfies:*

$$\delta_{k+1} < \frac{1}{\sqrt{k}+1} \tag{2.21}$$

then for k sparse signal $\mathbf{x}$, OMP will recover $\mathbf{x}$ from $\mathbf{y} = \mathbf{\Phi x}$ in k iterations.

This is the most relaxed requirement of *RIC* δ_{k+1} in the existing literatures. For detailed proof, please refer to (Mo and Shen, 2012). In (Mo and Shen, 2012), a counter example is given. A measurement dictionary with $\delta_{k+1} = \frac{1}{\sqrt{k}}$ is designed, but OMP algorithm fails. This means that there is no big space to improve the upper bound of *RIC* based on which OMP will definitely succeed.

In (Tropp, 2004), a sufficient condition for the success of OMP algorithm is given. Theorem 2.2 restates this sufficient condition.

Theorem 2.2. *(Tropp, 2004) (Exact Recovery for OMP): A sufficient condition for OMP to recover the sparse signal $\mathbf{x}$ from under determined linear measurement $\mathbf{y} = \mathbf{\Phi x}$ is that*

$$\max_{i \in \Gamma^c} \left| \mathbf{\Phi}_\Gamma^\dagger \varphi_i \right| < 1 \tag{2.22}$$

Theorem 2.2 states that OMP succeeds to recover sparse signal in compressive sensing in noiseless case so long as the exact condition holds true. In addition, (Tropp, 2004) also gives a sufficient condition for (2.22) based on cumulative coherence of measurement dictionary, which is stated in Theorem 2.3.

Theorem 2.3. *(Tropp, 2004) The Exact Recovery Condition in Theorem 2.2 holds*

whenever

$$\mu_1(k-1) + \mu_1(k) < 1 \tag{2.23}$$

Therefore, OMP algorithm is a successful algorithm for k sparse signal recovery whenever (2.23) holds true. Sufficient condition (2.23) is better than (2.22). That is because in (2.22), the index set Γ is needed. However, in signal recovery procedure, the index of correct atoms is not known beforehand. Therefore, (2.22) is only significant in theory. In contrast, sufficient condition (2.23) only involves cumulative coherence which is independent of the sparse signal. Thus, only based on measurement dictionary or more exactly based on cumulative coherence, we could evaluate whether OMP algorithm could recover k sparse signal successfully. However, please note that both conditions (2.22) and (2.23) are sufficient. For instance, when condition (2.23) is not met for certain measurement dictionary, OMP may still be able to recover k sparse signal. This is because the sufficient condition (2.22) and (2.23) are obtained based on the worst case analysis.

As stated in the last chapter, two parameters that are used to evaluate the quality of measurement dictionary, which are mutual coherence and RIC, are not independent. The following theorem gives the relationship of them.

Theorem 2.4. *There exists the following relationship between RIC and cumulative coherence for measurement dictionary in compressive sensing*

$$\delta_k \leq \mu_1(k-1) \tag{2.24}$$

Theorem 2.4 indicates that minimizing cumulative coherence would bound the

value of RIC. Although it is not realistic to calculate RIC of measurement dictionary especially when the size of measurement dictionary is pretty large, coherence or cumulative coherence of measurement dictionary is easy to calculate. The sufficient condition in Theorem 2.1 is with respect to RIC. Directly designing measurement dictionary with small RIC is not practical. However, based on Theorem 2.4, we can attempt to design a measurement dictionary with small coherence. (Elad (2007)), (Tropp *et al.* (2005)), (Li *et al.* (2013)) and (Rusu (2013)) aim to design measurement dictionary with small coherence or to design equi-angular tight frame. They also proved that with the designed measurement dictionary the performance of OMP algorithm improved a lot compared to Gaussian random measurement dictionary.

2.3 Incorporation of Sensing Dictionary in OMP Algorithm

Let matrix $\mathbf{\Psi} \in \mathbf{R}^{m \times n}$ be sensing dictionary, the size of which is the same as that of measurement dictionary. In OMP algorithm, if sensing dictionary is incorporated in the identification step in table 2.1, the identification step in OMP algorithm is changed as following:

$$i = \max_{1 \leq j \leq n} \left| \psi_j^T \mathbf{r} \right| \tag{2.25}$$

where $\psi_j \in \mathbf{R}^{m \times 1}$ is jth atom of sensing dictionary $\mathbf{\Psi}$. OMP algorithm with sensing dictionary is given in table 2.2. The only difference from OMP algorithm in table 2.1 is that the identification vector is calculated using sensing dictionary.

Input: measurement dictionary $\mathbf{\Phi}$, sensing dictionary $\mathbf{\Psi}$, measurement signal $\mathbf{y}$ and sparsity k.
repeat step 1 to 3 for k iterations
Step1(identification): $\mathbf{h} = \mathbf{\Psi}^T\mathbf{r}$, $i = \max_{1\leq j\leq n} \lvert\mathbf{h}(j)\rvert$
Step2(sparse signal reconstruction): $\Lambda = \Lambda \cup i$, $\tilde{\mathbf{x}} = \mathbf{\Phi}_\Lambda^\dagger \mathbf{y}$
Step3(residual update): $\mathbf{r} = \mathbf{r} - \Phi_\Lambda \tilde{\mathbf{x}}$
Output: reconstructed sparse signal with $\hat{\mathbf{x}}_\Lambda = \tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}_{\Lambda^c} = \mathbf{0}$.

Table 2.2: Orthogonal Matching Pursuit with sensing dictionary

As RIC of measurement dictionary, for sensing and measurement dictionaries, a similar value termed generalized RIC is defined.

Defination 2.3.1. *(Li* et al.*, 2011): If a pair of measurement and sensing dictionaries satisfies the following inequality for any k sparse signal $\mathbf{x}$*

$$(1-\delta_k')\left\|\mathbf{x}\right\|_2 \leq \left\|\mathbf{\Psi}^T\mathbf{\Phi}\mathbf{x}\right\|_2 \leq (1+\delta_k')\left\|\mathbf{x}\right\|_2 \tag{2.26}$$

the parameter δ_k' is termed kth order generalized restricted isometry constant(generalized RIC).

Like RIC for measurement dictionary, there is a relationship between the eigenvalue of type Gram matrix $\mathbf{\Psi}_\Gamma^T\mathbf{\Phi}_\Gamma$ and generalized RIC. For any index set Γ satisfying $|\Gamma| \leq k$, there exists the following inequality:

$$1-\delta_k' \leq \lambda_{min}\left(\mathbf{\Psi}_\Gamma^T\mathbf{\Phi}_\Gamma\right) \leq \lambda_{max}\left(\mathbf{\Psi}_\Gamma^T\mathbf{\Phi}_\Gamma\right) \leq 1+\delta_k' \tag{2.27}$$

where $\lambda_{min}\left(\mathbf{\Psi}_\Gamma^T\mathbf{\Phi}_\Gamma\right)$ and $\lambda_{max}\left(\mathbf{\Psi}_\Gamma^T\mathbf{\Phi}_\Gamma\right)$ denote the minimal and maximal eigenvalue of type Gram matrix $\mathbf{\Psi}_\Gamma^T\mathbf{\Phi}_\Gamma$ respectively. It can be seen from (2.27) that to calculate generalized RIC is computational difficult as the eigenvalue of total $\binom{n}{m}$ numbers of Gram matrix is required to be calculated.

Define identification vector as $\mathbf{h} = \mathbf{\Psi}^T\mathbf{r}$. Similar to (2.17), (2.18) and (2.19), when involved with sensing dictionary, the identification vector $\mathbf{h}$ and signal $\mathbf{x}_{T\setminus\Lambda}$ have the following relationship:

$$\left|\mathbf{h}\left(i\right) - \mathbf{x}_{\Gamma\setminus\Lambda}\left(i\right)\right| \leq \frac{\delta'_{k+1}}{1-\delta_{k+1}}\left\|\mathbf{x}_{\Gamma\setminus\Lambda}\right\|_2 \qquad i \in \Gamma\setminus\Lambda \tag{2.28}$$

$$\left|\mathbf{h}\left(i\right)\right| = 0 \qquad i \in \Lambda \tag{2.29}$$

$$\left|\mathbf{h}\left(i\right)\right| \leq \frac{\delta'_{k+1}}{1-\delta_{k+1}}\left\|\mathbf{x}_{\Gamma\setminus\Lambda}\right\|_2 \qquad i \in \Gamma^c \tag{2.30}$$

The analysis is similar to that for (2.17), (2.18) and (2.19). As can be seen from (2.28), (2.29) and (2.30), we could design a sensing dictionary to make generalized RIP δ'_{k+1} as small as possible so as to make the upper bound of perturbation small. As such, identification vector component $\mathbf{h}(i)$, $i \in \Gamma\setminus\Lambda$ approximates signal $\mathbf{x}_{\Gamma\setminus\Lambda}$ and the value of component of $\mathbf{h}(i)$, $i \in \Gamma^c$ is relatively small. OMP algorithm could select a right atom at each iteration easily.

Theorem 2.5 gives a sufficient condition based on which OMP algorithm involving sensing dictionary succeeds in recovering k sparse signal in compressive sensing.

Theorem 2.5. *Suppose that $k+1$th order generalized RIC of sensing and measurement dictionaries $\mathbf{\Psi}$ and $\mathbf{\Phi}$ satisfy the following inequality*

$$\delta'_{k+1} < \frac{1}{\sqrt{k}+1} \tag{2.31}$$

then for any k parse signal $\mathbf{x}$, OMP will recover $\mathbf{x}$ from $\mathbf{y} = \mathbf{\Phi x}$ in k iteration.

Proof. At the first iteration of OMP algorithm, based on generalized *RIC* (2.26) and identification vector (2.8), we obtain

$$\left(1 - \delta'_{k+1}\right) \|\mathbf{x}\|_2^2 \leq \mathbf{x}^T \mathbf{\Psi}^T \mathbf{\Phi x} = \mathbf{x}^T \mathbf{h} \tag{2.32}$$

$$\leq \|\mathbf{x}\|_1 \max_{i \in \Gamma} |\mathbf{h}(i)| \tag{2.33}$$

$$\leq \sqrt{k} \|\mathbf{x}\|_2 \max_{i \in \Gamma} |\mathbf{h}(i)| \tag{2.34}$$

this implies that

$$\max_{i \in \Gamma} |\mathbf{h}(i)| \geq \frac{\left(1 - \delta'_{k+1}\right) \|\mathbf{x}\|_2}{\sqrt{k}} \tag{2.35}$$

Moreover, for the component of identification vector corresponding to the wrong index, there is

$$|\mathbf{h}(i)| \leq \delta'_{k+1} \|\mathbf{x}\|_2 \qquad i \in \Gamma^c \tag{2.36}$$

Based on the above two inequalities, it is easy to verify that when the condition (2.31) is satisfied, OMP algorithm could select the right atom in the first iteration. After the first iteration, it is nontrivial to prove that OMP algorithm can select a right atom at the following each iteration when condition (2.31) is satisfied. The only modification should be made on (2.35) and (2.36) is to replace $\mathbf{x}$ with $\mathbf{x}_r$. □

The other parameters similar to coherence and cumulative coherence of measurement dictionary are mutual coherence and cumulative mutual coherence defined for sensing and measurement dictionaries as following:

Defination 2.3.2. *(Schnass and Vandergheynst, 2008): Let* $\mathbf{\Psi}$ *and* $\mathbf{\Phi}$ *are sensing and measurement dictionaries respectively. Mutual coherence and cumulative mutual coherence are defined as*

$$\mu' = \max_{1\leq i,j\leq n} \left|\psi_i^T \varphi_j\right| \tag{2.37}$$

$$\mu_1'(k) = \max_{|\Lambda|=k} \max_{1\leq i,j\leq n, i\notin\Lambda, j\in\Lambda} \sum_{\Lambda} \left|\psi_i^T \varphi_j\right| \tag{2.38}$$

where $\psi_i \in \mathbf{R}^{m\times 1}$ *is* i*th atom of sensing dictionary* $\mathbf{\Psi}$ *and* $\varphi_j \in \mathbf{R}^{m\times 1}$ *is* j*th atom of measurement dictionary* $\mathbf{\Phi}$*.*

With mutual coherence and cumulative mutual coherence defined in (2.37) and (2.38), sufficient condition for OMP algorithm is given in theorem 2.6.

Theorem 2.6. *(Schnass and Vandergheynst, 2008) In OMP algorithm, when sensing dictionary* $\mathbf{\Psi}$ *is involved, right atom will always be selected if*

$$\left\| \left(\mathbf{\Phi}_{\Gamma^c}^T \mathbf{\Psi}_{\Gamma}\right)^{-1} \mathbf{\Phi}_{\Gamma}^T \mathbf{\Psi}_{\Gamma^c} \right\|_{1,1} < 1 \tag{2.39}$$

which is always satisfied if

$$\mu_1'(k) + \mu_1'(k-1) < \beta \tag{2.40}$$

where $\beta = \min_{1\leq i\leq n} \left|\psi_i^T \varphi_i\right|$

Detailed proof can be found in (Schnass and Vandergheynst, 2008). As can be seen that, condition (2.39) is difficult to check whether it is satisfied as the index set Γ is not known beforehand. Moreover, in compressive sensing, index set is unknown

before hand and it is the main target of OMP algorithm. However, sufficient condition (2.40) only involves mutual coherence and cumulative mutual coherence, which are independent of the index of right atoms. Mutual coherence and cumulative mutual coherence are only determined by sensing and measurement dictionaries. Therefore, it is practical to determine whether the sufficient condition (2.40) is met based on cumulative mutual coherence of sensing and measurement dictionaries. Moreover, it can be seen from (2.40) that mutual cumulative coherence should be minimized by designing sensing dictionary with respect to the given measurement dictionary.

Minimization of cumulative mutual coherence also makes sense for generalized RIC. Like relationship between cumulative coherence and RIC for measurement dictionary, similar relationship exits between generalized RIC and cumulative mutual coherence. Theorem 2.7 states the relationship.

Theorem 2.7. *(Li* et al.*, 2011) $k-1$th order cumulative mutual coherence and kth order generalized RIC of sensing and measurement dictionaries have the following inequality:*

$$\delta'_k \leq \mu'_1 (k-1) \tag{2.41}$$

Theorem 2.7 indicates that minimization of $k-1$th order cumulative mutual coherence would bound the value of kth order generalized RIC. Generalized RIC with small value is favored by OMP algorithm according to (2.31). Moreover, cumulative mutual coherence with small value is also of special significance for OMP algorithm according to (2.40). Therefore, it is necessary to design sensing dictionary with respect to measurement dictionary to make cumulative mutual coherence as small as possible. There exists simple relationship between mutual coherence and cumulative

mutual coherence which is

$$\mu_1'(k) \leq k\mu' \tag{2.42}$$

Therefore, based on (2.42), we could design a sensing dictionary with respect to the given measurement dictionary with small mutual coherence to force the cumulative mutual coherence to be small. In the next chapter, sensing dictionary design algorithms will be proposed to design sensing dictionary with respect to the given measurement dictionary to reduce the value of mutual coherence.

Chapter 3

Sensing Dictionary Design Algorithms

In the previous chapter, sensing dictionary is introduced in OMP algorithm to improve the ability of OMP algorithm to select right atoms based on the identification vector. Designing sensing dictionary with respect measurement dictionary to force the mutual coherence or the mutual cumulative coherence as small as possible is favored by OMP algorithm. In this chapter, the existing sensing dictionary design algorithms will be introduced and novel sensing dictionary design methods will be given.

3.1 Existing Sensing Dictionary Design Algorithms

In (Schnass and Vandergheynst, 2008), sensing dictionary design algorithm using alternating projection method is proposed. Before introduction of the algorithm, two

sets are given here as

$$\mathtt{G} = \left\{\mathbf{G} = \mathbf{\Psi}^T\mathbf{\Phi}, \mathbf{\Psi} \in \mathbf{R}^{m\times n}, \mathbf{\Phi} \in \mathbf{R}^{m\times n}\right\} \tag{3.1}$$

$$\mathtt{H} = \left\{\mathbf{H} \in \mathbf{R}^{n\times n}, \mathbf{H}_{ii} = 1, \mathbf{H}_{ij} \leq \mu_w, for \quad i \neq j, 1 \leq i, j \leq n\right\} \tag{3.2}$$

where (3.1) is the set of type Gram matrix and (3.2) is the set comprising matrix whose diagonal entries are ones and the absolute value of off diagonal entries is less than or equal to Welch bound. $\mu_w = \sqrt{\frac{n-m}{m(n-1)}}$ is Welch bound for m by n matrix. $\mathbf{H}_{ij}$ denotes the component corresponding to the ith row and jth column of matrix $\mathbf{H}$. Based on the defined two matrix sets, sensing dictionary design problem could be written as following:

$$\min \|\mathbf{G} - \mathbf{H}\|_2 \qquad s.t \quad \mathbf{G} \in \mathtt{G}, \mathbf{H} \in \mathtt{H} \tag{3.3}$$

Problem (3.3) can be solved by alternating projection onto convex sets since both sets $\mathtt{G}$ and $\mathtt{H}$ are convex. In the alternating projection, the number of iteration is fixed and initialization is given as $\mathbf{G} = \mathbf{\Phi}^T\mathbf{\Phi}$, the following iterations are repeated until the maximum number of iteration is met:

1) find $\mathbf{H} \in \mathtt{H}$ to minimize $\|\mathbf{G} - \mathbf{H}\|_F$

2) find $\mathbf{G} \in \mathtt{G}$ to minimize $\|\mathbf{H} - \mathbf{G}\|_F$

where $\|\cdot\|_F$ is Frobenius norm of the input matrix.

In the first step, matrix $\mathbf{H}$ is to be calculated based on the minimization of Frobenius distance from type Gram matrix $\mathbf{G}$. The Frobenius norm can be written more

explicitly as

$$\min_{\mathbf{H}\in\mathbb{H}} \|\mathbf{G}-\mathbf{H}\|_F = \min_{\mathbf{H}\in\mathbb{H}} \left(\sum_{1\le i,j\le n} |\mathbf{G}_{ij}-\mathbf{H}_{ij}|^2 \right)^{1/2} \tag{3.4}$$

Based on (3.4), it is easy to get the expression of $\mathbf{H}$, which can be written as:

$$\mathbf{H}_{ij} = \begin{cases} 1 & if \quad i=j \\ \mathbf{G}_{ij} & if \quad |\mathbf{G}_{ij}| \le \mu_w \\ \mu_w sign\left(\mathbf{G}_{ij}\right) & if\, |\mathbf{G}_{ij}| > \mu_w \end{cases} \tag{3.5}$$

In the second step, given matrix $\mathbf{H}$, type Gram matrix whose Frobenius distance from $\mathbf{H}$ is minimal is required to calculate. For this, there exists a closed form solution to the problem in step two. The problem in step two can be rewritten as:

$$\begin{aligned} \min_{\mathbf{G}\in\mathbb{G}} \|\mathbf{G}-\mathbf{H}\|_F &= \min_{\mathbf{\Psi}} \left\|\mathbf{H}-\mathbf{\Psi}^T\mathbf{\Phi}\right\|_F \\ &= \min_{\mathbf{\Psi}} \left\|\mathbf{H}^T-\mathbf{\Phi}^T\mathbf{\Psi}\right\|_F \\ &= \min_{\mathbf{\Psi}} \left(\sum_{1\le i\le n} \left\|\mathbf{H}_{i\cdot}^T-\mathbf{\Phi}^T\psi_i\right\|_2^2 \right)^{1/2} \end{aligned} \tag{3.6}$$

where $\mathbf{H}_{i\cdot}$ denotes the ith row of matrix $\mathbf{H}$ and $\psi_i \in \mathbf{R}^{m\times 1}$ denotes the ith column of sensing dictionary $\mathbf{\Psi}$. It can be seen easily from (3.6) that the minimization problem can be decomposed into n independent minimization problems and each with its solution as

$$\psi_i = \left(\mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}\mathbf{H}_{i\cdot}^T \tag{3.7}$$

This could easily lead to the closed form of sensing dictionary as

$$\mathbf{\Psi} = \left(\mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}\mathbf{H}^T \tag{3.8}$$

The solutions to problems in both step one and two are obtained. The two steps iterate a certain number of iterations and a corresponding sensing dictionary will be designed. Alternating projection method minimizes the Frobenius distance between the set of type Gram matrix and the set of given matrix. As such, for the designed sensing dictionary, the absolute value of off diagonal entries of type Gram matrix is near the Welch bound. However, it can not guarantee that its diagonal entries are of unit value.

Another kind of sensing dictionary design algorithm is dependent on both measurement dictionary and measurement signal (Huang *et al.*, 2011). In (Huang *et al.*, 2011), besides measurement dictionary, measurement signal is used as a weighting coefficient to minimize the objection function. The minimization problem can be formulated as

$$\min_{\mathbf{\Psi}\in\mathbf{R}^{m\times n}} \left\|\mathbf{\Psi}^T\mathbf{\Phi}\mathbf{W}\right\|_F^2 \qquad s.t. \quad \psi_i^T\varphi_i = 1, 1 \le i \le n \tag{3.9}$$

where $\mathbf{W} = diag\{|\mathbf{h}|\}$ is a weighting matrix which provides posterior knowledge. In (3.9), the objective function can be further written as

$$\left\|\mathbf{\Psi}^T\mathbf{\Phi}\mathbf{W}\right\|_F^2 = \sum_{i=1}^{n}\left\|\mathbf{W}\mathbf{\Phi}^T\psi_i\right\|_2^2 \tag{3.10}$$

Therefore, the minimization problem (3.9) can be factorized as n sub minimization

problem with the ith one as

$$\min_{\psi_i \in \mathbf{R}^{m\times 1}} \left\| \boldsymbol{W}\boldsymbol{\Phi}^T\psi_i \right\|_2^2 \qquad s.t. \quad \psi_i^T\varphi_i = 1 \tag{3.11}$$

The Lagrangian for (3.11) is

$$L\left(\psi_i, \lambda\right) = \left\| \mathbf{W}\boldsymbol{\Phi}^T\psi_i \right\|_2^2 + \lambda\left(\psi_i^T\varphi_i - 1\right) \tag{3.12}$$

It is nontrivial to obtain the solution when minimizing (3.12):

$$\psi_i = \frac{\mathbf{R}^{-1}\varphi_i}{\varphi_i^T\mathbf{R}^{-1}\varphi_i} \quad where \quad \mathbf{R} = \boldsymbol{\Phi}\mathbf{W}^2\boldsymbol{\Phi}^T \tag{3.13}$$

It can been seen from (3.13) that a closed form solution exists. The constructed sensing dictionary could reduce local cumulative mutual coherence (local cumulative mutual coherence is defined in (Huang *et al.*, 2011)). The performance of OMP algorithm improves significantly when incorporating the designed sensing dictionary. However, one fatal drawback of this algorithm is that it depends on measurement signal. In practice, for each new measurement signal, a corresponding sensing dictionary is required to be designed, which is computationally expansive.

3.2 Sensing Dictionary Design Based on Linear Programming

In this section, a novel sensing dictionary design algorithm is proposed. The aim of the new algorithm is to minimize the mutual coherence and the corresponding

problem could be formulated as linear programming problem.

According to the definition of sensing dictionary in the last chapter, if all the diagonal entries of type Gram matrix are one, then the maximum absolute value of all the off diagonal entries of type Gram matrix represents mutual coherence. Like the method in (Schnass and Vandergheynst, 2008) and (Huang *et al.*, 2011), we intend to construct sensing dictionary atom by atom or column by column. To design the ith column ($1 \leq i \leq n$) ψ_i, we intend to solve the following optimization problem

$$\min_{\psi_i \in \mathbf{R}^{m\times 1}} \left\| \psi_i^T \mathbf{\Phi}_\Lambda \right\|_\infty \qquad s.t. \quad \psi_i^T \varphi_i = 1 \tag{3.14}$$

where $\Lambda = \{1, 2 \cdots n\} \setminus i$.

It can be seen from (3.14) that to design ith atom of sensing dictionary, the constraint is that its inner product with the ith atom of measurement dictionary is equal to one. The maximum absolute value of its inner product with atoms other than the ith one of measurement dictionary is minimized. According to the definition of mutual coherence, mutual coherence is reduced when the optimization problem (3.14) is solved.

The optimization problem (3.14) is equivalent to linear programming as following:

$$\min_{\psi_i \in \mathbf{R}^{m\times 1}, \tau \in \mathbf{R}} \tau \qquad s.t. \quad \psi_i^T \varphi_i = 1, \quad \left| \psi_i^T \varphi_j \right| \leq \tau, \quad j \in \{1, 2 \cdots n\} \setminus i \tag{3.15}$$

Problem (3.15) could be solved using linear programming. (3.15) can be written in standard linear programming format as

$$\min_{\tilde{\mathbf{x}} \in \mathbf{R}^{(m+1)\times 1}} \mathbf{c}^T \tilde{\mathbf{x}} \qquad s.t. \quad \mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{0}_{2(n-1)\times 1}, \quad \mathbf{d}^T \tilde{\mathbf{x}} = 1 \tag{3.16}$$

where $\tilde{\mathbf{x}} = \begin{bmatrix} \psi_i^T & \tau \end{bmatrix}^T$, $\mathbf{c} = \begin{bmatrix} \underbrace{0, 0 \cdots 0}_{m}, 1 \end{bmatrix}^T$, $\mathbf{d} = \begin{bmatrix} \varphi_i^T, 0 \end{bmatrix}^T$ and

$$\mathbf{A} = \begin{bmatrix} \varphi_1 & -\varphi_1 & \cdots & \varphi_{i-1} & -\varphi_{i-1} & \varphi_{i+1} & -\varphi_{i+1} & \cdots & \varphi_n & -\varphi_n \\ -1 & -1 & \cdots & -1 & -1 & -1 & -1 & \cdots & -1 & -1 \end{bmatrix}^T$$

When implemented in Matlab, we could use function `linprog.m` to solve the linear programming problem (3.16). The problem is convex and global optimal solution could be found definitely. Usually, primal dual interior point algorithm is implemented to solve linear programming problem. Logarithmic barrier function is used for inequality equations and incorporating the logarithmic barrier functions into the objective function, which could be written as

$$\min_{\tilde{\mathbf{x}} \in \mathbf{R}^{(m+1)\times 1}} \mathbf{c}^T \tilde{\mathbf{x}} + \sum_{j=1}^{2(n-1)} \left(-\frac{1}{t} \right) ln \left(\mathbf{A}_{j\cdot} \tilde{\mathbf{x}} \right) \quad s.t. \quad \mathbf{d}^T \tilde{\mathbf{x}} = 1 \tag{3.17}$$

where $\mathbf{A}_{j\cdot}$ denotes the jth row of matrix $\mathbf{A}$. The modified *Karush Kuhn Tucker* (KKT) condition could be written as

$$\mathbf{c} + \mathbf{A}^T \lambda + \mathbf{d}^T \nu = 0 \tag{3.18}$$

$$diag\left\{\lambda\right\} \mathbf{A}\tilde{\mathbf{x}} = -\frac{1}{t}\mathbf{1} \tag{3.19}$$

$$\mathbf{d}^T \tilde{\mathbf{x}} = 1 \tag{3.20}$$

where $\lambda \in \mathbf{R}^{2(n-1)\times 1}$ and $\nu \in \mathbf{R}$ are dual variables for inequality and equality equations respectively. $\mathbf{1}$ is a $2\left(2n-1\right)$ by 1 vector with all the element equal to 1. If $\tilde{\mathbf{x}}$, λ and ν satisfy modified KKT equations (3.18), (3.19) and (3.20), $\tilde{\mathbf{x}}$ is primal feasible

and λ, ν are dual feasible.

Define dual residual $\mathbf{r}_{dual}$, primal residual $\mathbf{r}_{pri}$ and centrality residual $\mathbf{r}_{cent}$ as

$$\mathbf{r}_{dual} = \mathbf{c} + \mathbf{A}^T\lambda + \mathbf{d}^T\nu \tag{3.21}$$

$$\mathbf{r}_{pri} = \mathbf{d}^T\tilde{\mathbf{x}} - 1 \tag{3.22}$$

$$\mathbf{r}_{cent} = -diag\{\lambda\}\,\mathbf{A}\tilde{\mathbf{x}} - \frac{1}{t}\mathbf{1} \tag{3.23}$$

Using Newton step to solve the KKT equations for fixed t. Assume current point $(\tilde{\mathbf{x}},\lambda,\nu)$ satisfies $\mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{0}$ and $\lambda \geq \mathbf{0}$, the Newton step $\Delta\mathbf{x}$, $\Delta\lambda$ and $\Delta\nu$ can be computed from the following equation:

$$\begin{bmatrix} \mathbf{0} & \mathbf{A}^T & \mathbf{b} \\ -diag\{\lambda\}\,\mathbf{A} & -diag\{\mathbf{A}\tilde{\mathbf{x}}\} & \mathbf{0} \\ \mathbf{b}^T & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_{dual} \\ \mathbf{r}_{cent} \\ \mathbf{r}_{pri} \end{bmatrix} \tag{3.24}$$

Equation (3.24) is a linear equation with total number of $m+2+2(n-1)$ equalities. The linear equation (3.24) can be solved by standard solution for linear equation such as conjugate gradient algorithm.

When the Newton step is obtained by solving equation (3.24), a step length could be determined by traditional backtracking line search method. For each fixed t, Newton step and the step length are calculated iteratively until the norm of three residual is below a given threshold. After that, the value of parameter t is decreased. The detailed procedure of primal dual interior point algorithm is not illustrated here. (Boyd and Vandenberghe, 2004) gives more related explanations.

3.3 The Proposed Sensing Dictionary Design Algorithm

In this part, a novel sensing dictionary design algorithm is given. As given in (3.14), the aim here is to reduce the maximal absolute value of inner product between the sensing dictionary atom and measurement dictionary while keeping the linear constrain satisfied.

In this algorithm, each atom of sensing dictionary is designed separately. When designing one specific atom of sensing dictionary, the algorithm is as follows: in each iteration, a new direction vector is calculated. When the value of the atom of sensing dictionary changes along the direction vector, the maximal absolute value of its inner product with measurement dictionary decreases and the linear constraint is satisfied. As an illustration, take the construction of the first sensing dictionary atom ψ_1 for example. Initialize ψ_1 as the first atom of measurement dictionary such as $\psi_1 = \varphi_1$. At the first iteration, we firstly search for the atom in measurement dictionary whose absolute value of inner product with ψ_1 is maximal

$$i_1 = arg \max_{j=1\cdots n, j\neq 1} \left|\psi_1^T \varphi_j\right| \tag{3.25}$$

We want to change atom ψ_1 with a direction $\mathbf{d}$ to minimize the maximal absolute value of its inner product with measurement dictionary $\mathbf{\Phi}$. It can be formulated as

an optimization problem as

$$\begin{aligned} &\min_{\mathbf{d}\in\mathbf{R}^{m\times 1}} \left|(\psi_1+\mathbf{d})^T\varphi_{i_1}\right| \\ s.t. \quad &(\psi_1+\mathbf{d})^T\varphi_1 = 1 \\ &\left|(\psi_1+\mathbf{d})^T\varphi_{i_1}\right| \geq \left|(\psi_1+\mathbf{d})^T\varphi_j\right|, j\in\{2,3,\cdots n\} \end{aligned} \tag{3.26}$$

We do not aim to solve the optimization problem (3.26) directly. We intend to solve it analytically. Define the index set $\Gamma = \{i_1\}$. As can be seen from (3.25), if we change the atom ψ_1 along the vector $-\varphi_{i_1}$, the maximal absolute value of the inner product with φ_{1_i} can be reduced. However, changing atom ψ_1 along the direction $-\varphi_{i_1}$ breaks the equality constraint. As a tradeoff, we change the atom ψ_1 along a direction that is orthogonal to φ_1 and along with direction $-\varphi_{i_1}$. This direction can be calculated as the complementary projection of $-\varphi_{i_1}$ on φ_1. This direction vector can be calculated as

$$\mathbf{d}_1 = -\varphi_{i_1} - \varphi_1\left(\varphi_1^T\varphi_1\right)^{-1}\varphi_1^T\left(-\varphi_{i_1}\right) \tag{3.27}$$

It is easy to prove that direction vector $\mathbf{d}_1$ is orthogonal to the atom φ_1, which is as following:

$$\begin{aligned} \varphi_1^T\mathbf{d}_1 &= \varphi_1^T\left(-\varphi_{i_1} + \varphi_1\left(\varphi_1^T\varphi_1\right)^{-1}\varphi_1^T\varphi_{i_1}\right) \\ &= -\varphi_1^T\varphi_{i_1} + \varphi_1^T\varphi_1\left(\varphi_1^T\varphi_1\right)^{-1}\varphi_1^T\varphi_{i_1} \\ &= 0 \end{aligned} \tag{3.28}$$

Another factor we should consider is the sign of vector $\mathbf{d}_1$. If we want to lower the maximum magnitude of inner product along direction $\mathbf{d}_1$, vector $\mathbf{d}_1$ should be

modified as

$$\mathbf{d}_1' = -\frac{sign\left(\psi_1^T \varphi_{i_1}\right)}{\mathbf{d}_1^T \varphi_{i_1}} \mathbf{d}_1 \tag{3.29}$$

where $sign(\cdot)$ denotes the sign of the input value.

As such, it is easy to check that inner product $\mathbf{d}_1'^T \varphi_{i_1}$ and $\psi_1^T \varphi_{i_1}$ are opposite in sign, which could be verified as following:

$$\mathbf{d}_1'^T \varphi_{i_1} = -\frac{sign\left(\psi_1^T \varphi_{i_1}\right)}{\mathbf{d}_1^T \varphi_{i_1}} \mathbf{d}_1^T \varphi_{i_1} = -sign\left(\psi_1^T \varphi_{i_1}\right) \tag{3.30}$$

We can change vector ψ_1 along the direction $\mathbf{d}_1'$ with nonnegative step length α, which can be written as

$$\psi_1 + \alpha \mathbf{d}_1' \tag{3.31}$$

The inner product of the new vector and φ_{i_1} can be expressed as

$$\begin{aligned} \left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_{i_1} &= \psi_1^T \varphi_{i_1} + \alpha \mathbf{d}_1'^T \varphi_{i_1} \\ &= \psi_1^T \varphi_{i_1} - \alpha sign\left(\psi_1^T \varphi_{i_1}\right) \\ &= \left|\psi_1^T \varphi_{i_1}\right| sign\left(\psi_1^T \varphi_{i_1}\right) - \alpha sign\left(\psi_1^T \varphi_{i_1}\right) \\ &= \left[\left|\psi_1^T \varphi_{i_1}\right| - \alpha\right] sign\left(\psi_1^T \varphi_{i_1}\right) \end{aligned} \tag{3.32}$$

The absolute value of the inner product is

$$\left|\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_{i_1}\right| = \left|\left|\psi_1^T \varphi_{i_1}\right| - \alpha\right| \tag{3.33}$$

It can been seen that if $\alpha \leq \left|\psi_1^T \varphi_{i_1}\right|$, comparing to the absolute value of $\psi_1^T \varphi_{i_1}$, the value in (3.33) decreases by α. Based on this observation, it is easy to obtain the following result:

$$\left|\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_{i_1}\right| \leq \left|\psi_1^T \varphi_{i_1}\right| \qquad if \quad 0 \leq \alpha \leq \left|\psi_1^T \varphi_{i_1}\right| \tag{3.34}$$

Intuitively, when the value of α is $\alpha = 0^+$, which means that it is a small positive value, based on (3.30), it is estimated that the absolute value of $\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_{i_1}$ will decrease. If we increase the value of α further, the absolute value of certain inner product $\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_j$, $j \in \{2, 3, \cdots n\} \setminus \Gamma$ would increase. We can never expect that changing the value of α to make all the inner product $\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_j$, $j \in \{2, 3, \cdots n\}$ become zero. As in case that it happens, ψ_1 is orthogonal to all the vector φ_j, $j \in \{2, 3, \cdots n\} \setminus \Gamma$. This means that atoms φ_j, $j \in \{2, 3, \cdots n\} \setminus \Gamma$ could span a subspace and the atom φ_1 is not in this space. Otherwise, equality constrained is violated. In practice, it is not true as measurement dictionary satisfies $Kruskal$ rank or $krank$, which means that any m atoms from measurement dictionary are independent. It is impossible to form a subspace using atom φ_j, $j \in \{2, 3, \cdots n\} \setminus \Gamma$ and φ_1 is not in the subspace.

We can increase the value of α until the value of $\left|\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_{i_1}\right|$ becomes equal to the absolute value of inner product of $\psi_1 + \alpha \mathbf{d}_1'$ with certain atom in measurement dictionary. For atom φ_j, it means that

$$\left|\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_j\right| = \left|\left(\psi_1 + \alpha \mathbf{d}_1'\right)^T \varphi_{i_1}\right| \tag{3.35}$$

Solving this equation, we could obtain the value of α as

$$\alpha = -\frac{\psi_1^T(\varphi_{i_1} + \varphi_j)}{\mathbf{d}_1'(\varphi_{i_1} + \varphi_j)} \quad or \quad \alpha = -\frac{\psi_1^T(\varphi_{i_1} - \varphi_j)}{\mathbf{d}_1'(\varphi_{i_1} - \varphi_j)} \tag{3.36}$$

For each atom φ_j, $j \in \{2, 3, \cdots n\} \setminus \Gamma$, we could solve equation (3.35) and get two solutions with explicit form as (3.36). Therefore, there exists $2(n-1)$ value for α. We should choose the value of α as the smallest positive value among them, such as

$$\alpha = \min_{j=\{2,3,\cdots n\}\setminus\Gamma} \left\{ \left[-\frac{\psi_1^T(\varphi_{i_1} \pm \varphi_j)}{\mathbf{d}_1'^T(\varphi_{i_1} \pm \varphi_j)}, 0 \right]_+ \right\} \tag{3.37}$$

where operator $[\cdot]_+$ is defined as following:

$$[x, 0]_+ = \begin{cases} x & if \quad x > 0 \\ 0 & otherwise \end{cases} \tag{3.38}$$

Let φ_{i_2} be the atom that corresponds to expression that gives the value of α in (3.37), which is

$$i_2 = \arg_{j\in\{2,3\cdots n\}\setminus\Gamma} \left\{ \alpha = \left[-\frac{\psi_1^T(\varphi_{i_1} + \varphi_j)}{\mathbf{d}_1'^T(\varphi_{i_1} + \varphi_j)}, -\frac{\psi_1^T(\varphi_{i_1} - \varphi_j)}{\mathbf{d}_1'^T(\varphi_{i_1} - \varphi_j)} \right]_+ \right\} \tag{3.39}$$

Renew the index set as $\Gamma = \Gamma \cup i_2$. According to (3.35), α satisfies the following equation

$$\left| (\psi_1 + \alpha \mathbf{d}_1')^T \varphi_{i_2} \right| = \left| (\psi_1 + \alpha \mathbf{d}_1')^T \varphi_{i_1} \right| \tag{3.40}$$

If we update the atom ψ_1 as the following

$$\psi_1 \leftarrow \psi_1 + \alpha \mathbf{d}'_1 \tag{3.41}$$

then the absolute value of $\psi_1^T \varphi_{i_1}$ and $\psi_1^T \varphi_{i_2}$ is equal and maximal among all the inner product $\psi_1^T \varphi_j$, $j \in \{2, 3, \cdots n\}$.

At the second iteration, firstly we could obtain another direction vector as following:

$$\mathbf{d}_2 = -\varphi_{i_2} - \varphi_1 \left(\varphi_1^T \varphi_1\right)^{-1} \varphi_1^T \left(-\varphi_{i_2}\right) - \mathbf{d}_1 \left(\mathbf{d}_1^T \mathbf{d}_1\right)^{-1} \mathbf{d}_1^T \left(-\varphi_{i_2}\right) \tag{3.42}$$

which is the complementary projection of $-\varphi_{i_2}$ on the space spanned by $\mathbf{d}_1$ and φ_1, as $\mathbf{d}_1$ and φ_1 are orthogonal. It is easy to verify the following equalities:

$$\mathbf{d}_2^T \varphi_1 = 0, \quad \mathbf{d}_2^T \mathbf{d}_1 = 0 \tag{3.43}$$

which means that directional vector $\mathbf{d}_2$ is orthogonal to φ_1 and $\mathbf{d}_1$.

Now, we have two directional vector which are $\mathbf{d'}_1$ and $\mathbf{d}_2$. Both of them are orthogonal to φ_1, so we could change ψ_1 along the direction spanned by them without violating the equality constraint. In detail, changing ψ_1 along $\mathbf{d}'_1$ could reduce the value of $\psi_1^T \varphi_{i_1}$ and changing ψ_1 along $\mathbf{d}_2$ could reduce the value of $\psi_1^T \varphi_{i_2}$. As $\mathbf{d'}_1$ and $\mathbf{d}_2$ are orthogonal, we could construct a new directional vector $\mathbf{d}'_2$ as

$$\mathbf{d}'_2 = \mathbf{d}'_1 + c_2 \mathbf{d}_2 \tag{3.44}$$

where $c_2 \in \mathbf{R}$ is an unknown coefficient.

If we change ψ_1 along the direction $\mathbf{d}'_2$ with step length α, the atom could be written as $\psi_1 + \alpha\mathbf{d}'_2$. The inner product of this atom with φ_{i_1} is

$$\begin{aligned}\left(\psi_1 + \alpha\mathbf{d}'_2\right)^T \varphi_{i_1} &= \left(\psi_1 + \alpha\mathbf{d}'_1 + \alpha c_2 \mathbf{d}_2\right)^T \varphi_{i_1} \\ &= \left(\psi_1 + \alpha\mathbf{d}'_1\right)^T \varphi_{i_1} \\ &= \left[\left|\psi_1^T \varphi_{i_1}\right| - \alpha\right] sign\left(\psi_1^T \varphi_{i_1}\right)\end{aligned} \tag{3.45}$$

where the second equality is based on the orthogonality of $\mathbf{d}_2$ and φ_{i_1}, the third equality is based on (3.32). Similar to (3.32), when we change ψ_1 along directional vector $\mathbf{d}'_2$ with step length α, if the value of α is no larger than $\left|\psi_1^T \varphi_{i_1}\right|$, the magnitude of its inner product with φ_{i_1} will decrease by α, which is independent of the coefficient c_2. In the same way, we wish that when changing ψ_1 in the direction $\mathbf{d}'_2$ with step length α, the magnitude of its inner product with ϕ_{i_2} decreases by α. This could expressed as

$$\begin{aligned}\left(\psi_1 + \alpha\mathbf{d}'_2\right)^T \varphi_{i_2} &= \left(\psi_1 + \alpha\mathbf{d}'_1 + \alpha c_2 \mathbf{d}_2\right)^T \varphi_{i_2} \\ &= \left[\left|\psi_1^T \varphi_{i_2}\right| - \alpha\right] sign\left(\psi_1^T \varphi_{i_2}\right)\end{aligned} \tag{3.46}$$

Based on (3.46), we could get the value of coefficient c_2 as

$$c_2 = -\frac{sign\left(\psi_1^T \varphi_{i_2}\right) + \mathbf{d'}_1^T \varphi_{i_2}}{\mathbf{d}_2^T \varphi_{i_2}} \tag{3.47}$$

We can get the expression of directional vector $\mathbf{d'}_2$ by substituting c_2 in (3.44) with the value in (3.47). As such, both of the magnitude of $\left(\psi_1 + \alpha\mathbf{d'}_2\right)\varphi_{i_1}$ and $\left(\psi_1 + \alpha\mathbf{d'}_2\right)\varphi_{i_2}$ change linearly with α with the same speed. The value of α could be

determined in the same way at the first iteration as (3.35), which could be written as

$$\alpha = \min_{j=\{2,3,\cdots n\}\backslash\Gamma} \left\{ \left[-\frac{\psi_1^T \left(\varphi_{i_1} \pm \varphi_j\right)}{\mathbf{d}_2'^T \left(\varphi_{i_1} \pm \varphi_j\right)}, 0 \right]_+ \right\} \tag{3.48}$$

After $t-1$th $(t-1 < m-1)$ iteration, the directional vector is $\mathbf{d}'_{t-1}$ and the selected index is i_{t-1}. The index set could be renewed as $\Gamma = \Gamma \cup i_{t-1}$. At tth iteration, a direction vector is obtained as

$$\mathbf{d}_t = -\varphi_{i_t} - \varphi_1 \left(\varphi_1^T \varphi_1\right)^{-1} \varphi_1^T \left(-\varphi_{i_t}\right) - \sum_{j=1}^{t-1} \mathbf{d}_j \left(\mathbf{d}_j^T \mathbf{d}_j\right)^{-1} \mathbf{d}_j^T \left(-\varphi_{i_t}\right) \tag{3.49}$$

which means that $\mathbf{d}_t$ is complementary projection of $-\varphi_{i_t}$ on the subspace spanned by φ_1 and $\mathbf{d}_j$,$(j = 1, 2, \cdots, t-1)$. A new directional vector $\mathbf{d}'_t$ could be written as

$$\mathbf{d}'_t = \mathbf{d}'_{t-1} + c_t \mathbf{d}_t \tag{3.50}$$

As (3.47), c_t could be explicitly written as

$$c_t = -\frac{sign\left(\psi_1^T \varphi_{i_t}\right) + \mathbf{d}'^T_{t-1} \varphi_{i_t}}{\mathbf{d}_t^T \varphi_{i_t}} \tag{3.51}$$

and the step length could be determined the same way as in the first and second steps, which is

$$\alpha = \min_{j=\{2,3,\cdots n\}\backslash\Gamma} \left\{ \left[-\frac{\psi_1^T \left(\varphi_{i_{t-1}} \pm \varphi_j\right)}{\mathbf{d}_t'^T \left(\varphi_{i_{t-1}} \pm \varphi_j\right)}, 0 \right]_+ \right\} \tag{3.52}$$

Note that the maximum number of iterations is $m-1$. As at tth iteration, a directional vector $\mathbf{d}_t$ is calculated as complementary projection on the subspace

spanned by φ_1 and $\mathbf{d}_j$,$(j = 1, 2, \cdots, t-1)$. When the number of iteration is m, the upper mentioned spanned subspace is full subspace and the complementary projection of any vector is a zero vector. This is due to the krank property of measurement dictionary.

Table 3.1 gives the algorithm proposed in this section. The algorithm is termed sensing dictionary design algorithm. There are in total two loops in this algorithm. In the outer loop, each atom of sensing dictionary is iterated and in the inner loop one atom of sensing dictionary is designed.

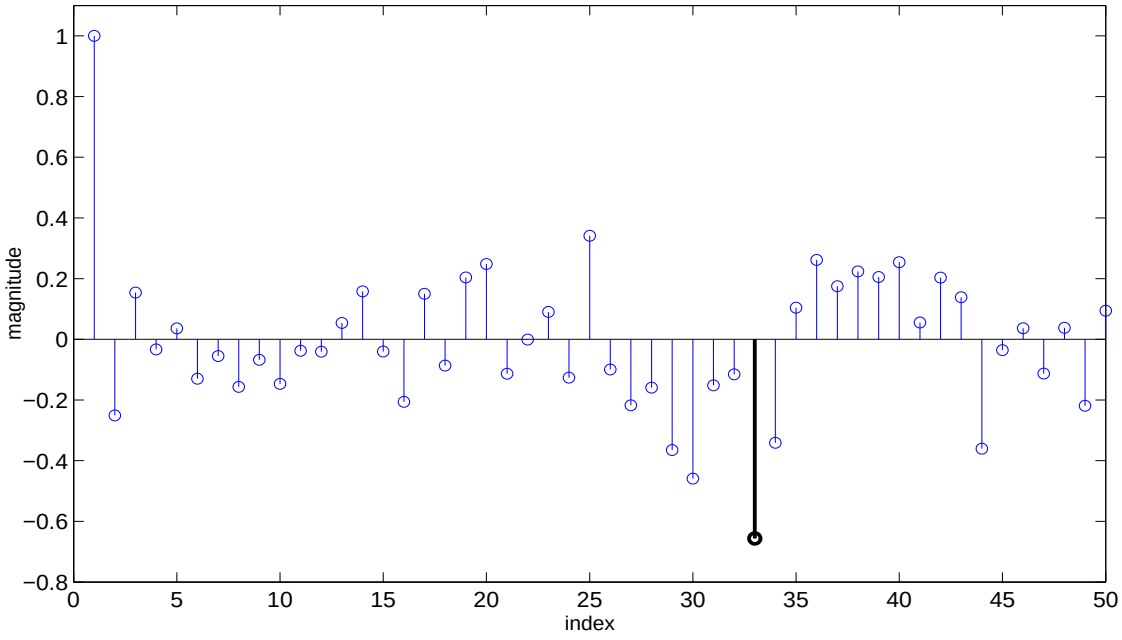

Figure 3.1: Inner product of atom ψ_1 and measurement dictionary Φ in initialization step

Here, we give an example which aims to design the first atom of sensing dictionary $\psi_\mathbf{1}$ when the measurement dictionary Φ is given. Each component of measurement dictionary is generated as random Gaussian value with zero mean and unit variance. After that, each column of measurement dictionary is normalized to make each atom

Input: measurement dictionary $\mathbf{\Phi}$.

for i = 1 : n

Initialization: ψ_i=φ_i. $i_1 = arg \quad max \left|\psi_i^T \varphi_j\right|$ where $j = \{1, 2, \cdots n\} \setminus \{i\}$. $\Gamma = \{i_1\}$. Calculating $\mathbf{d}_1$ and $\mathbf{d'}_1$ using (3.27) and (3.29). Calculating α as in (3.37). Renewing atom as $\psi_i = \psi_i + \alpha \mathbf{d}'_1$

for t = 2 : m-1

Step1: $i_t = \arg_{j=\{1,2,\cdots,n\}\setminus\{i\cup\Gamma\}} \left\{\alpha = \left[-\frac{\psi_i^T\left(\varphi_{i_{t-1}} \pm \varphi_j\right)}{\mathbf{d'}_{t-1}^T\left(\varphi_{i_{t-1}} \pm \varphi_j\right)}\right]_+\right\}$, $\Gamma = \Gamma \cup \{i_t\}$.

Step2: calculating $\mathbf{d}_t$, c_t, $\mathbf{d}'_t$ and α as (3.49), (3.51) and (3.50) and (3.52).

Step3: $\psi_i = \psi_i + \alpha \mathbf{d}'_t$

end

end

Output: sensing dictionary $\mathbf{\Psi}$

Table 3.1: Sensing Dictionary Design Algorithm

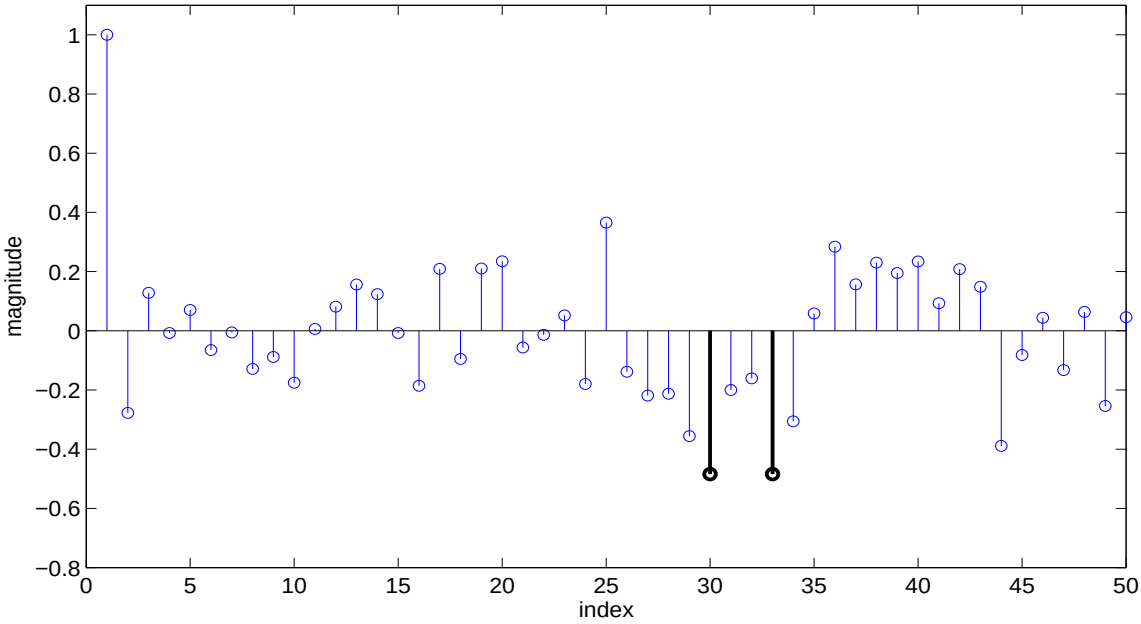

Figure 3.2: Inner product of atom ψ_1 and measurement dictionary Φ after the first iteration

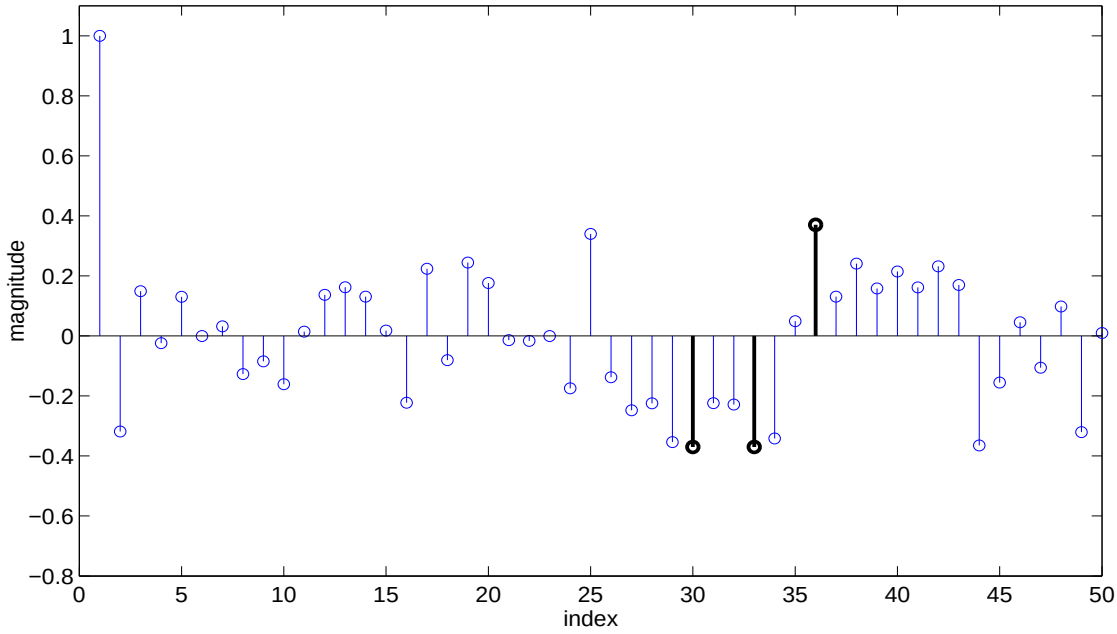

Figure 3.3: Inner product of atom ψ_1 and measurement dictionary Φ after the second iteration

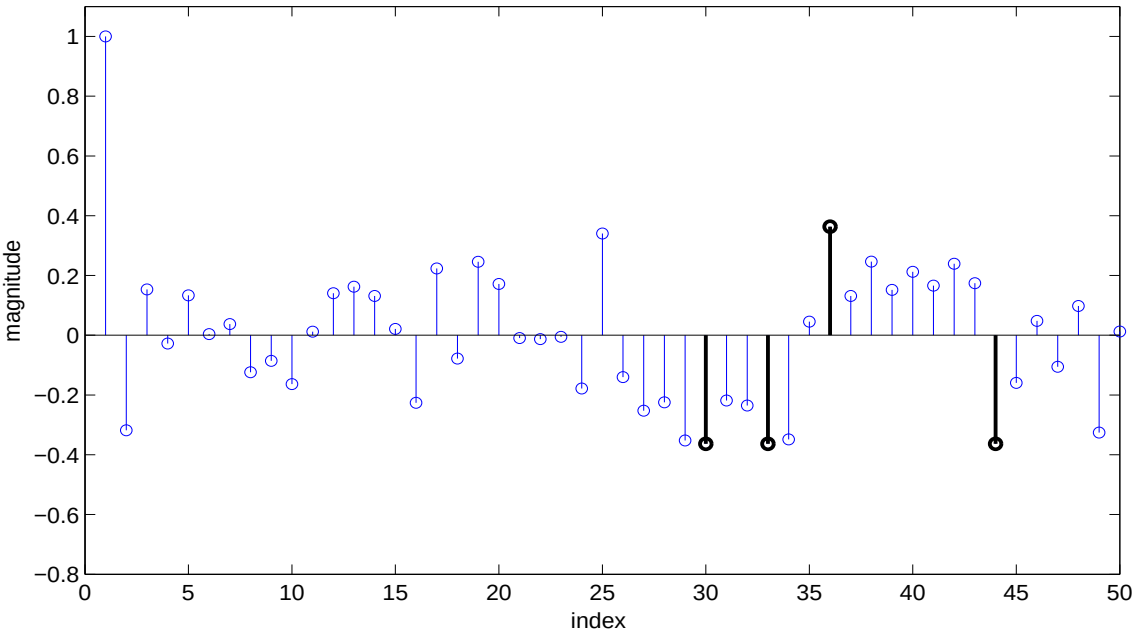

Figure 3.4: Inner product of atom ψ_i and measurement dictionary Φ after the third iteration

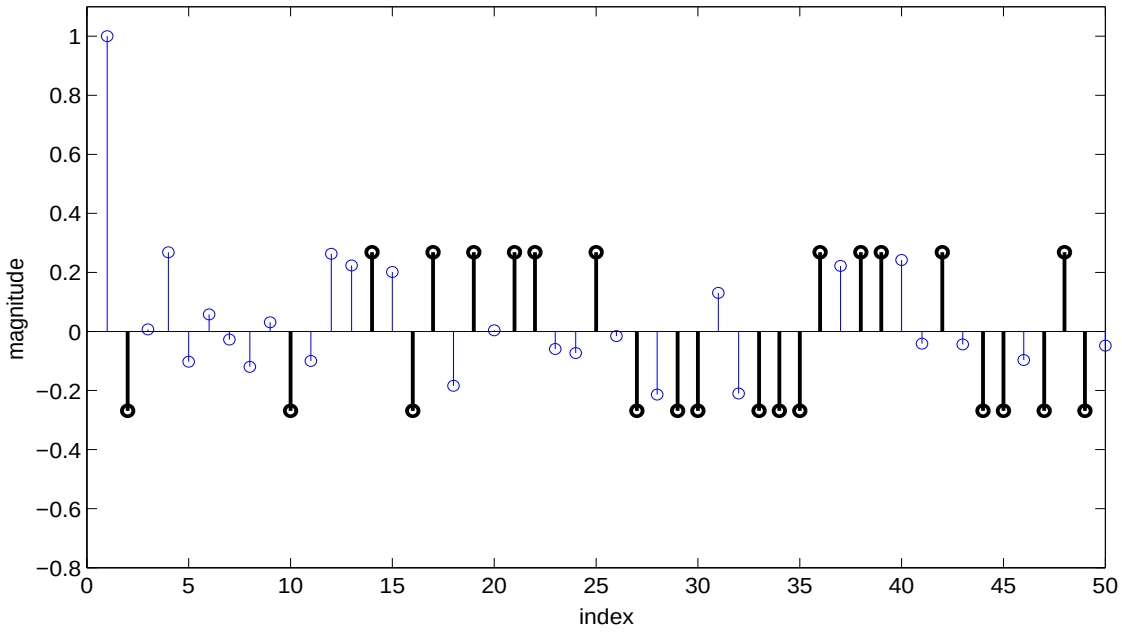

Figure 3.5: Inner product of atom ψ_i and measurement dictionary after the $m-2$th iteration

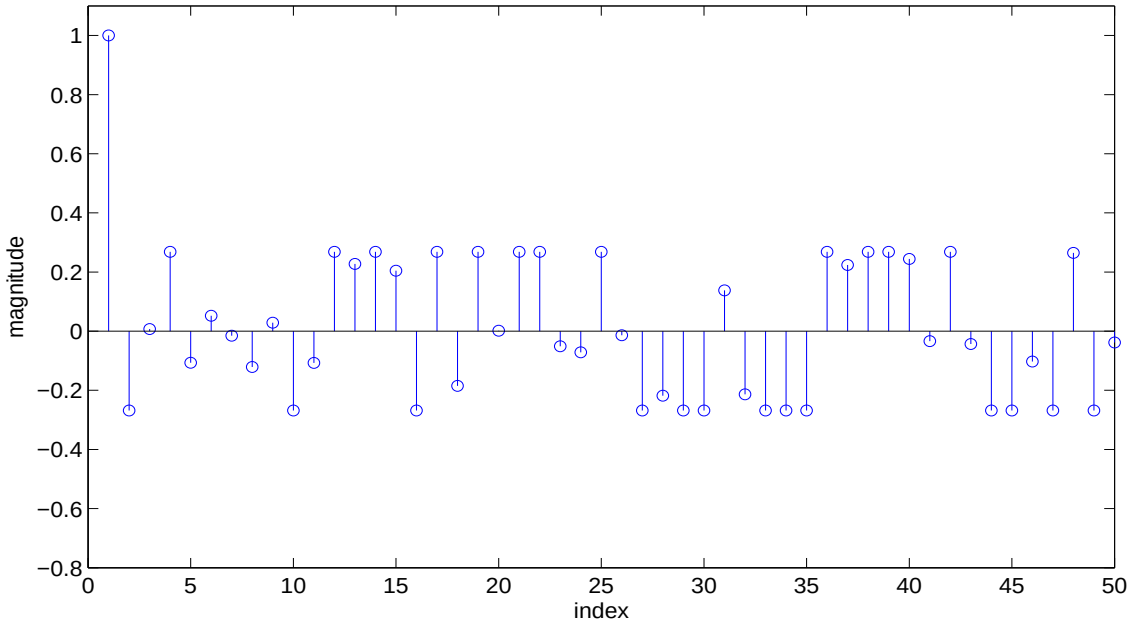

Figure 3.6: Inner product of atom ψ_i and measurement dictionary. Atom ψ_i is designed by linear programming based algorithm

with unit two norm. The size of measurement dictionary is 25 by 50. Figure 3.1 gives the inner product of ψ_1 and measurement dictionary $\mathbf{\Phi}$. This is also the initialization step of sensing dictionary design algorithm. It can be clearly seen that the inner product $\psi_1^T\phi_{33}$ is of largest magnitude among all the inner product $\psi_1^T\phi_j$, $j = 2, 3, \cdots, 50$. Moreover, $\psi_1^T\phi_1 = 1$ means the linear constraint is satisfied. Figure 3.2 shows the inner product of ψ_1 and Φ after the first iteration. We could see that both of the inner products $\psi_1^T\phi_{33}$ and $\psi_1\phi_{30}$ are of largest magnitude and are reduced compared to figure 3.1. In other words, the largest magnitude of inner product $\psi_1^T\phi_j$, $j = 2, 3, \cdots, 50$ decreases. Figure 3.3 and 3.4 are the inner product of ψ_1 and Φ after second and third iterations. We could see that the largest magnitude of the inner product decreases as the number of iteration grows. After the last iteration, the inner product is illustrated in figure 3.5. Comparing to figure 3.1, the largest

magnitude decreases and there exists 24 components of the inner product that are of equal magnitude which is largest.

Sensing dictionary atom ψ_1 can be also designed by linear programming based method given at the previous section. The inner product of the designed sensing dictionary atom ψ_1 and measurement dictionary Φ is shown in figure 3.6. Comparing figure 3.5 and figure 3.6, it is clear that the two inner products are identical, which means that the performance of sensing dictionary design algorithm and linear programming are the same in this example.

Compared with linear programming method, there is no guarantee that sensing dictionary design algorithm could design a sensing dictionary that is globally optimal. However, simulation manifests that sensing dictionary design algorithm performs as well as linear programming method.

3.4 Complexity Analysis

In this section, the complexity of linear programming method and sensing dictionary design algorithm is briefly analyzed.

For linear programming based method, sensing dictionary is designed atom by atom. When designing each atom, primal dual interior point algorithm is used. In primal dual interior point algorithm, for fixed parameter t, a series of Newton steps is implemented to calculate the Newton step and the corresponding step length is calculated by backtracking line search method. Each Newton step is obtained by solving linear programming as in (3.24) with complexity $O\left((m+2n)^3\right)$. Denote the N_{newton} and N_t being the number of Newton step iterations for each fixed parameter t and the number of parameter t. Although for each fixed parameter t the number

of Newton steps varies, we use N_{newton} to represent the number of Newton step iterations for simplicity. Moreover, in practice, the number of Newton step iterations for each fixed t is of the same order of magnitude. Thus the use of N_{newton} is relatively reasonable. Based on the above analysis and assumption, the computational complexity of primal dual interior point algorithm to design one sensing dictionary atom is $O\left(N_t N_{newton}\left(m+2n\right)^3\right)$. In compressive sensing when the length of signal $\mathbf{x}$ is large, the computation of sensing dictionary is burdensome.

In sensing dictionary design algorithm, when designing one sensing dictionary atom, at jth iteration the computational complexity of complementary projection (3.49) is $O\left(jm\right)$ and the computational complexity of calculation of α in (3.52) is $O\left(m\left(n-j\right)\right)$. These two steps are the most computational consuming in this algorithm. In total, at jth step, the computational complexity is $O\left(mn\right)$. The number of iterations is known to be $m-1$. Thus, computational complexity of sensing dictionary design algorithm to design one atom is $O\left(nm^2\right)$.

Comparing the complexity of the two algorithms above, we could see that sensing dictionary design algorithm is computationally simpler than linear programming based method.

Chapter 4

Simulation and Comparison

In this part, the performance of the proposed sensing dictionary design algorithm and linear programming based sensing dictionary design algorithm along with the algorithm in (Schnass and Vandergheynst, 2008) will be compared.

4.1 Comparison of Sensing Dictionaries

Each component of measurement dictionary is generated as random Gaussian value with zero mean and unit variance. After that, each column of measurement dictionary is normalized to make two norm of each atom of measurement dictionary with unit value. Three algorithms are used to design sensing dictionary, which are the proposed sensing dictionary design algorithm, linear programming based method and algorithm in (Schnass and Vandergheynst, 2008).

Figure 4.1 compares the cumulative mutual coherence with sensing dictionary constructed by the method in (Schnass and Vandergheynst, 2008) and two algorithms proposed in this thesis. The size of the dictionary is 128 by 256. When using sensing

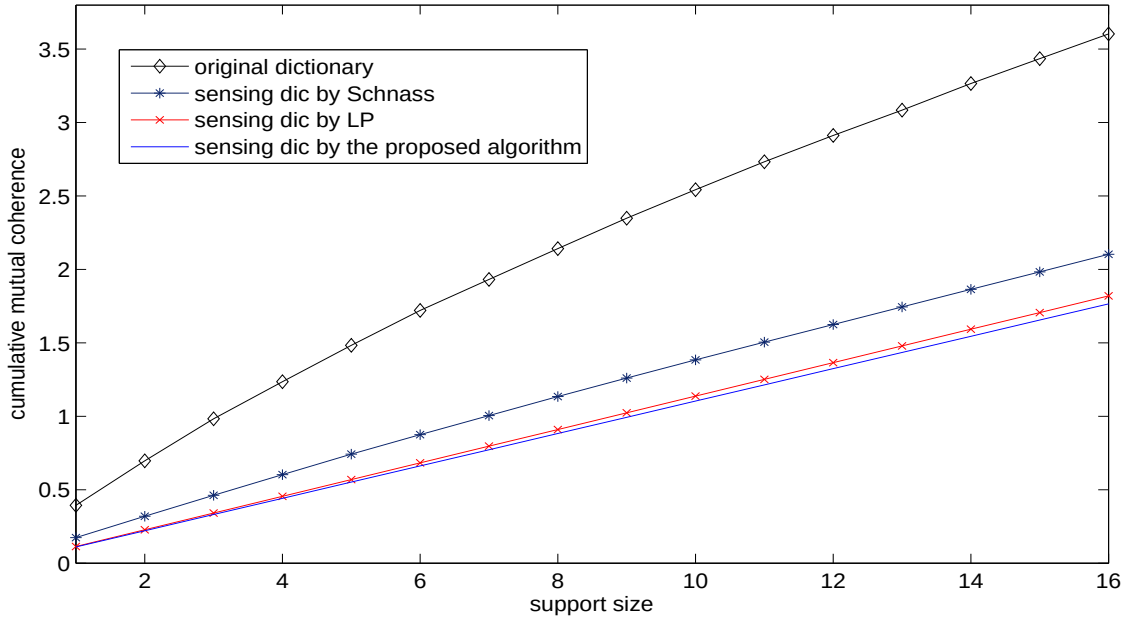

Figure 4.1: Comparison of cumulative mutual coherence

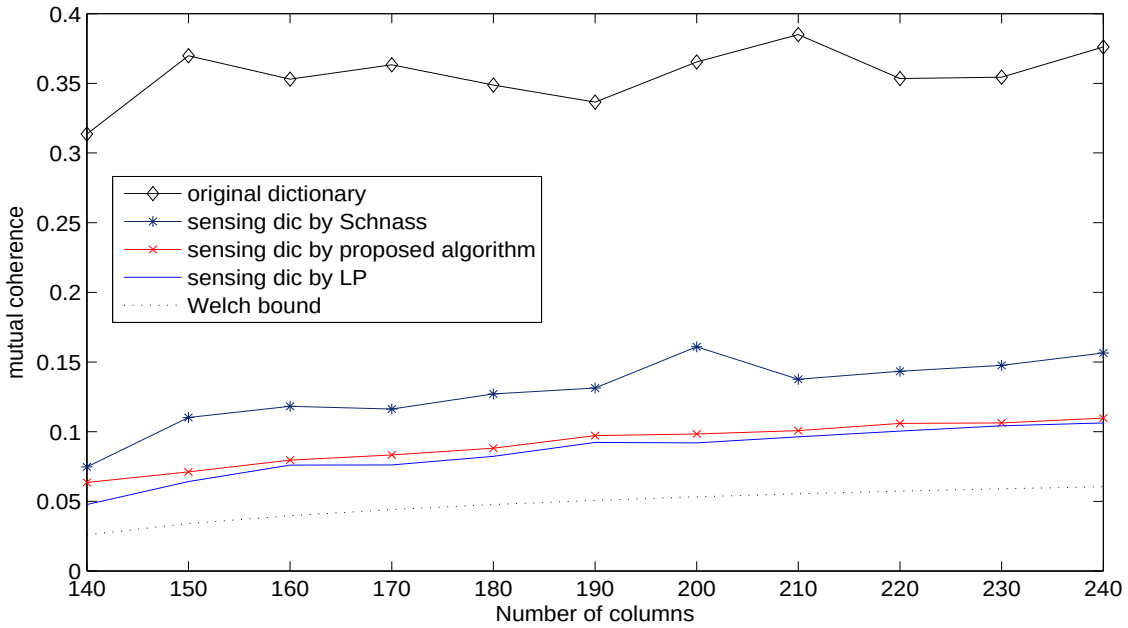

Figure 4.2: Comparison of mutual coherence with various dictionary size

dictionary designed by the linear programming based algorithm or sensing dictionary design algorithm, cumulative mutual coherence are smaller than that of sensing dictionary by (Schnass and Vandergheynst, 2008). Moreover, for the sensing dictionary constructed either by linear programming based algorithm or sensing dictionary desig algorithm, the cumulative mutual coherence is relatively identical, although the cumulative mutual coherence by the latter algorithm is larger than the former one.

Figure 4.2 compares mutual coherence with sensing dictionary designed by the upper mentioned three algorithms. The number of row of the dictionary is 128 and the number of columns varies from 140 to 240 with interval equal to 10. Welch bound is also illustrated as a comparison. As can be clearly seen, with sensing dictionary designed by linear programming based method or the proposed sensing dictionary design algorithm, the mutual coherence is smaller than that with sensing dictionary given by the method in (Schnass and Vandergheynst, 2008).

As stated in the last chapter, linear programming based method is computational burdensome while the proposed sensing dictionary design algorithm is rather computationally simple. Although computational complexity is analyzed in the last chapter and algorithm running time using Matlab does not accurately represent the computational complexity, the latter gives an intuitional realization of computational complexity. In figure 4.3, the running time of the two algorithms proposed is illustrated. The number of columns is 256 and the number of rows varies from 100 to 200 with interval equal to 10. Two algorithms are running under Matlab 7.11.0(R2010b) in laptop with Intel i7-2620M CUP 2.70 and 2.69 GHz and 8.00G RAM. For both the algorithm, running time grows with the increased number of rows. As can be seen from figure 4.3 that the running time of the proposed sensing dictionary designing

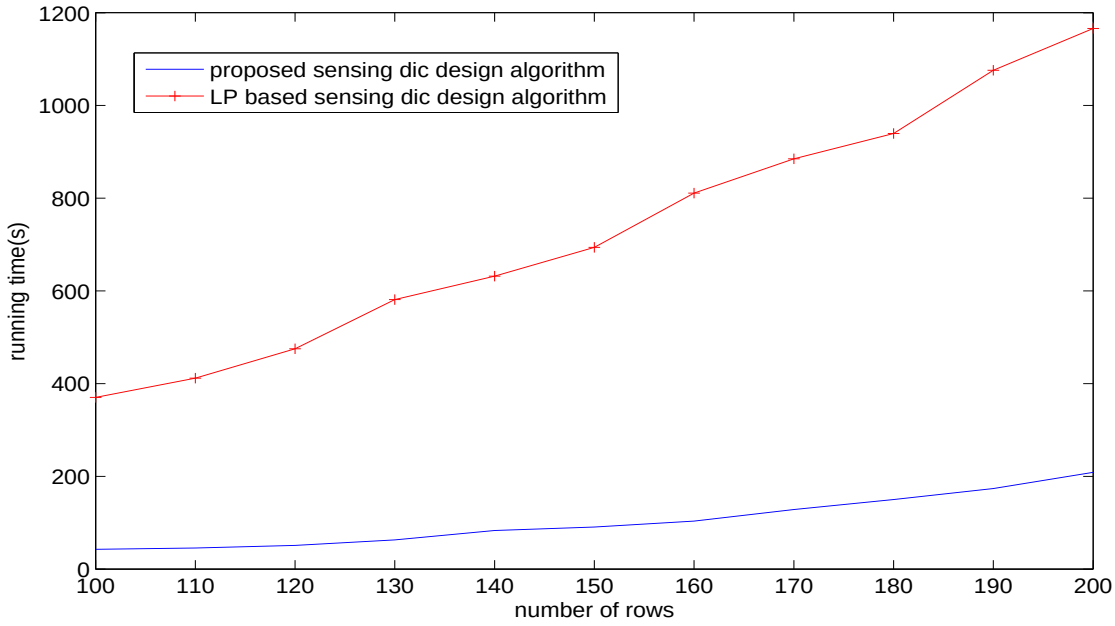

Figure 4.3: Comparison of running time of linear programming based algorithm and the proposed sensing dictionary design algorithm

algorithm is 5 to 10 times less than linear programming based algorithm.

4.2 Comparison of OMP Algorithm Incorporated with Sensing Dictionary

In this section, the performance of OMP algorithm is compared when using sensing dictionary designed by different kinds of algorithms. The size of dictionary is 128 by 256. The number of nonzero component of sparse signal varies from 1 to 79 and for each sparsity level the experiment is run 500 times. Sparse signal recovery is claimed to be successful if the support of the recovered signal and that of the original sparse signal are identical.

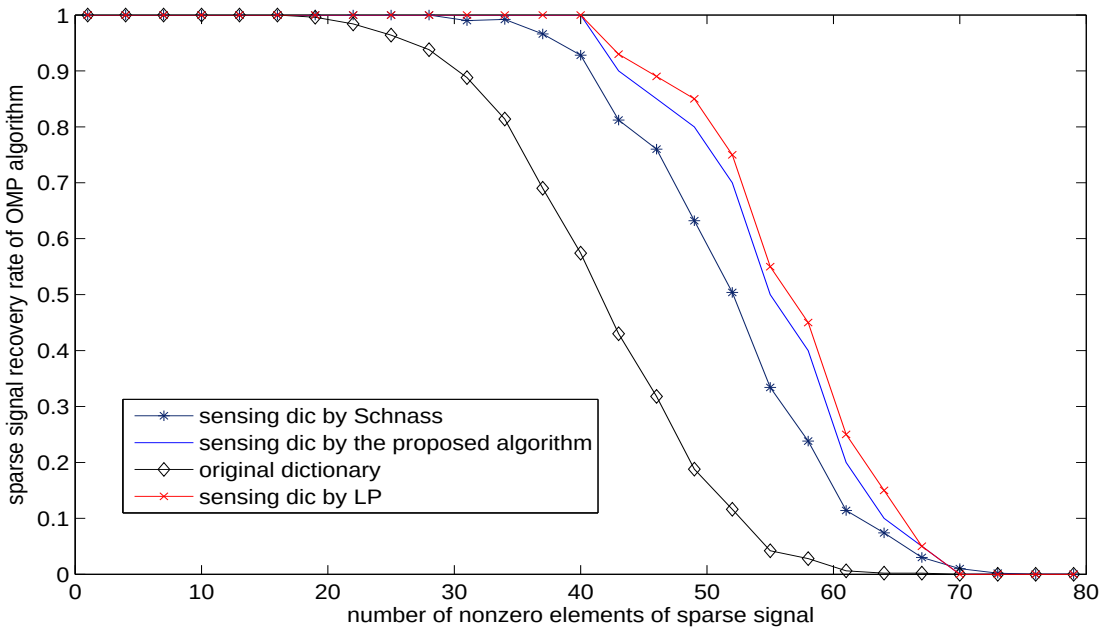

Figure 4.4: Performance of OMP algorithm when the nonzero component of sparse signal is Gaussian distributed

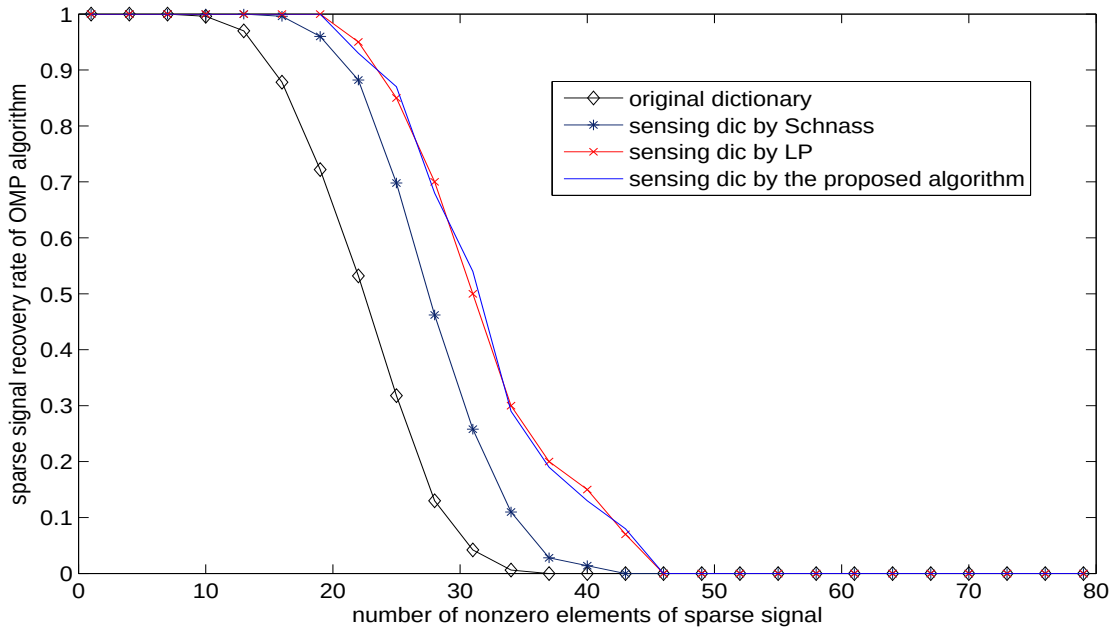

Figure 4.5: Performance of OMP algorithm when the nonzero component of sparse signal is equal to one

In figure 4.4, the nonzero component of sparse signal is with standard Gaussian distribution and in figure 4.5 the nonzero component of sparse signal is of unit value. From figure 4.4 and 4.5, we could see that the performance of OMP algorithm improves when using sensing dictionary. Especially, the sensing dictionary designed by linear programming based algorithm or the proposed sensing dictionary design algorithm makes OMP algorithm perform better than using sensing dictionary by the algorithm in (Schnass and Vandergheynst, 2008). Moreover, when using sensing dictionary designed by the two proposed algorithm, the performance of OMP algorithm is identical.

Chapter 5

Conclusion

In this thesis, two sensing dictionary design algorithms are proposed to improve the performance of OMP algorithm when recovering sparse signal in compressive sensing. The aim of sensing dictionary design is to make the mutual coherence and cumulative mutual coherence small, which benefits OMP algorithm.

In the first algorithm, sensing dictionary atom design problem is formulated as an optimization problem with linear equalities and linear inequalities constrictions. This problem can be transformed as a linear programming problem and primal-dual interior point algorithm is utilized to solve it. Global optimal point is guaranteed to exist because of the convexity of the problem. However, high computational complexity is a major obstacle.

In the second algorithm, each sensing dictionary atom is designed to make the maximal magnitude of its inner product with measurement dictionary as small as possible. Compared with the first one, the second algorithm carries low computational complexity while global optimal solution is not guaranteed.

Simulation verifies that both of the proposed algorithms could design sensing

dictionary with respect to the given measurement dictionary to make the mutual and cumulative mutual coherence small. The performance of OMP algorithm also improves when incorporating the sensing dictionary designed by the two proposed algorithms.

Further research direction includes derivation of theoretical sufficient conditions under which the sensing dictionaries designed by the two proposed algorithms in this thesis are identical. Another idea is to minimize the cumulative mutual coherence directly in the optimization problem, which may give a designed sensing dictionary with even smaller cumulative mutual coherence.

Bibliography

Boyd, S. P. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Candes, E. J. and Tao, T. (2005). Decoding by linear programming. *Information Theory, IEEE Transactions on*, **51**(12), 4203–4215.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, **20**(1), 33–61.

Davenport.M.A and Wakin.M.B (2010). Analysis of orthogonal matching pursuit approach via restricted isometry property. *Information Theory, IEEE Transactions on*, **56**(9), 4395–4401.

Donoho, D. L. (2006). Compressed sensing. *Information Theory, IEEE Transactions on*, **52**(4), 1289–1306.

Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., and Baraniuk, R. G. (2008). Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE*, **25**(2), 83–91.

Elad, M. (2007). Optimized projections for compressed sensing. *Signal Processing, IEEE Transactions on*, **55**(12), 5695–5702.

Gorodnitsky, I. F. and Rao, B. D. (1997). Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm. *Signal Processing, IEEE Transactions on*, **45**(3), 600–616.

Huang, A., Gui, G., Wan, Q., and Mehbodniya, A. (2011). A re-weighted algorithm for designing data dependent sensing dictionary. *International Journal of the Physical Sciences*, **6**(3), 386–390.

Huggins, P. S. and Zucker, S. W. (2007). Greedy basis pursuit. *Signal Processing, IEEE Transactions on*, **55**(7), 3760–3772.

Ji, S., Xue, Y., and Carin, L. (2008). Bayesian compressive sensing. *Signal Processing, IEEE Transactions on*, **56**(6), 2346–2356.

Li, B., Shen, Y., and Li, J. (2011). Dictionaries construction using alternating projection method in compressive sensing. *Signal Processing Letters, IEEE*, **18**(11), 663–666.

Li, G., Zhu, Z., Yang, D., Chang, L., and Bai, H. (2013). On projection matrix optimization for compressive sensing systems.

Malioutov, D., Çetin, M., and Willsky, A. S. (2005). A sparse signal reconstruction perspective for source localization with sensor arrays. *Signal Processing, IEEE Transactions on*, **53**(8), 3010–3022.

Mo, Q. and Shen, Y. (2012). A remark on the restricted isometry property in orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, **58**(6), 3654–3656.

Mourad, N. and Reilly, J. P. (2010). Minimizing nonconvex functions for sparse vector reconstruction. *Signal Processing, IEEE Transactions on*, **58**(7), 3485–3496.

Rusu, C. (2013). Design of incoherent frames via convex optimization.

Schnass, K. and Vandergheynst, P. (2008). Dictionary preconditioning for greedy algorithms. *Signal Processing, IEEE Transactions on*, **56**(5), 1994–2002.

Tropp, J. A. (2004). Greed is good: Algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, **50**(10), 2231–2242.

Tropp, J. A. and Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, **53**(12), 4655–4666.

Tropp, J. A., Dhillon, I. S., Heath, R. W., and Strohmer, T. (2005). Designing structured tight frames via an alternating projection method. *Information Theory, IEEE Transactions on*, **51**(1), 188–209.