

Chapter 2

Compressed Sensing

In this chapter compressed sensing is introduced in more details. Gelfand’s width, which is a pure mathematical concept with close connection to CS, is introduced in Sect. 2.1. CS in restricted isometry perspective is considered in Sect. 2.2. Section 2.3 covers a review on spherical section property, which will be used in the next Chapter.

2.1 Gelfand’s Width

Some mathematical ideas that are used in CS originally came from the Harmonic Analysis literature. In this section we introduce Gelfand’s width and show how it is connected with CS theory. Let $S \subset \mathbb{R}^n$ and $m < n \in \mathbb{N}$. Assume \mathbb{R}^n is equipped with l_p -norm.

Definition 2 *Gelfand’s width for this set is defined as:*

$$d^m(S)_p = \inf_K \sup\{\|\mathbf{x}\|_p | \mathbf{x} \in S \cap K\} = \inf_K \sup_{\mathbf{x} \in S \cap K} \|\mathbf{x}\|_p, p \geq 1, \quad (2.1)$$

where infimum is taken over all $n - m$ dimensional subspace K of \mathbb{R}^n . Assume S be bounded such that:

$$\begin{aligned} \forall s \in S : -s \in S \\ \exists a \in \mathbb{R}^n : S + S \subset aS \end{aligned} \quad (2.2)$$

For instance if $S = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x}\| < 1\}$, then assuming $a = 2$, this set satisfies (2.2). Now assume we sample elements of S with a sampling matrix $\Phi \in \mathbb{R}^{m \times n}$. Also let D be an operator (possibly nonlinear) which is used for reconstructions:

$$\begin{aligned} \mathbf{y} &= \Phi \mathbf{x} \\ \hat{\mathbf{x}} &= D(\mathbf{y}) \end{aligned} \quad (2.3)$$

The error reconstruction in this sampling/reconstruction system over the set S would be:

$$E(S, \Phi, D) = \sup_{\mathbf{x} \in S} \|\mathbf{x} - \hat{\mathbf{x}}\|_p = \sup_{\mathbf{x} \in S} \|\mathbf{x} - D(\Phi\mathbf{x})\|_p \quad (2.4)$$

We are interested in finding a (Φ, D) pair such that $E(S, \Phi, D)$ is minimized. The best possible performance in this framework is given by:

$$E(S) = \inf_{\Phi, D} E(S, \Phi, D). \quad (2.5)$$

As we know $\dim(\text{Null}(\Phi)) = n - m$ and thus $\text{null}(\Phi)$ is a $n - m$ dimensional subspace of \mathbb{R}^n and can be considered an instance of K in the definition of Gelfand's width of S :

$$d^m(S)_p = \inf_K \sup_{\mathbf{x} \in S \cap K} \|\mathbf{x}\|_p \leq \sup_{\mathbf{x} \in S \cap \text{null}(\Phi)} \|\mathbf{x}\|_p. \quad (2.6)$$

On the other hand:

$$\forall \mathbf{x} \in S \cap \text{null}(\Phi) : D(\mathbf{y}) = D(\Phi\mathbf{x}) = D(0) = D(-\Phi\mathbf{x}) \quad (2.7)$$

Now note:

$$\|\mathbf{x} - D(\Phi\mathbf{x})\|_p + \|\mathbf{x} - D(-\Phi\mathbf{x})\|_p \geq \|\mathbf{x} - D(0) - \mathbf{x} + D(0)\|_p = 2\|\mathbf{x}\|_p \rightarrow \quad (2.8)$$

$$\|\mathbf{x} - D(0)\|_p \geq \|\mathbf{x}\|_p \quad \text{or} \quad \|\mathbf{x} - D(0)\|_p \geq \|\mathbf{x}\|_p.$$

Thus for any $\mathbf{x} \in S \cap \text{null}(\Phi)$, there exists an element $\mathbf{x}' \in S \cap \text{null}(\Phi)$ such that: $\|\mathbf{x}' - D(\Phi\mathbf{x}')\|_p \geq \|\mathbf{x}\|_p$. Consider this fact:

$$E(S, \Phi, D) \geq \sup_{\mathbf{x} \in S \cap K} \|\mathbf{x} - D(\Phi\mathbf{x})\|_p \geq \sup_{\mathbf{x} \in S \cap \text{null}(\Phi)} \|\mathbf{x} - D(\Phi\mathbf{x})\|_p \geq \sup_{\mathbf{x} \in S \cap \text{null}(\Phi)} \|\mathbf{x}\|_p. \quad (2.9)$$

From (2.6) and taking infimum on (2.9) one can conclude:

$$E(S)_p \geq d^m(S)_p. \quad (2.10)$$

Now assume $K \subset \mathbb{R}^n$ with $\dim(K) = n - m$. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ be a basis for orthogonal complement of K (K^\perp). Form the sampling matrix $\Phi = [\mathbf{v}_1, \dots, \mathbf{v}_m]^T$. Also we define the reconstruction operator D as follows:

$$D(\mathbf{u}) = \begin{cases} \mathbf{a} & \text{if } \mathbf{u} \in \Phi S \\ \mathbf{b} & \text{if } \mathbf{u} \notin \Phi S \end{cases} \quad (2.11)$$

where $\mathbf{a} \in S$ is arbitrary such that $\mathbf{u} = \Phi\mathbf{a}$ and \mathbf{b} is a randomly chosen vector in S . With these assumptions on (Φ, D) we calculate $E(S)_p$. Let $\mathbf{x} \in S$:

$$\Phi(\mathbf{x} - D(\Phi\mathbf{x})) = \Phi\mathbf{x} - \Phi D(\Phi\mathbf{x}) = \Phi\mathbf{x} - \Phi\mathbf{x} = 0, \quad (2.12)$$

which yields $\mathbf{x} - D(\Phi\mathbf{x}) \in \text{null}(\Phi) \equiv K$. Also from (2.2):

$$\exists a \in \mathbb{R} : \frac{\mathbf{x} - D(\Phi\mathbf{x})}{a} \in S \rightarrow \frac{\mathbf{x} - D(\Phi\mathbf{x})}{a} \in S \cap K. \quad (2.13)$$

Consequently:

$$\begin{aligned} E(S, \Phi, D)_p &= a \sup_{\mathbf{x} \in S} \left\| \frac{\mathbf{x} - D(\Phi\mathbf{x})}{a} \right\|_p \leq a \sup_{\mathbf{x} \in S \cap K} \|\mathbf{x}\|_p \rightarrow \\ \inf_{\Phi, D} E(S, \Phi, D)_p &\leq a \inf_K \sup_{\mathbf{x} \in S \cap K} \|\mathbf{x}\|_p \rightarrow E(S)_p \leq ad^m(S)_p. \end{aligned} \quad (2.14)$$

Overall from (2.10) and (2.14):

$$d^m(S)_p \leq E(S)_p \leq ad^m(S)_p. \quad (2.15)$$

This is an important result and shows how the reconstruction error over the set S is related to Gelfand's width of S . In other words, the best reconstruction performance in CS is bounded by Gelfand's width. Unfortunately finding Gelfand's width of a set in the general case is an open problem and only for special instances of S , such as unit ball, Solutions have been found. Advances in this area provide a strong mathematical imbed for CS theory.

The central question is what Φ, D pair would satisfy the bounds given by Gelfand's width? Independently, in [1, 2] sufficient condition on the sensing matrix was provided. The authors introduced the concept of restricted isometry property (RIP) and used this concept to provide theorems for unique and stable source reconstruction and prove the CS theorems. They showed that the random sensing matrices with i.i.d. Gaussian or Bernoulli entries satisfy the required conditions and efficient decoding D , can be accomplished by linear programming as in (1.4) (this reconstruction method has been provided before through empirical approach).

2.2 Restricted Isometry Property and Coherence

The classical theory of CS [1, 2] uses the concept of RIP. As discussed in Chap. 1, if the source is k -sparse, then if any combination of $2k$ columns of A is linearly independent, then the solution of (1.2) would be unique. Having this in mind, the restricted isometry property (RIP) is defined as follows:

Definition 3 *Restricted Isometry Property*

We say an arbitrary matrix A , satisfies RIP of order k with constant $0 \leq \delta_k < 1$, if for all k -sparse vectors \mathbf{x} :

$$1 - \delta_k \leq \frac{\|A\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \leq 1 + \delta_k. \quad (2.16)$$

This means that k -sparse sources not only will not lay in the null-space of A , but also will have a distance δ_k with this space. This condition is stronger compared to linear independency of any $2k$ columns of A and in return is also stable towards noise. In other words it means that all sub-matrices of A with at most k columns are well-conditioned. The constant $0 \leq \delta_k < 1$ measures closeness of the sensing operator to an orthonormal system. From discussions in Chap. 1, one concludes if A satisfies RIP with $0 \leq \delta_{2k} < 1$ then the solution of (1.2) is unique and can be recovered through solving (1.3). But for practical applications equivalence of the solutions of (1.3) and (1.4) is essential.

Historically CS results are developed using RIP and over the time the conditions and bounds on theorems are improved. The following two theorems are two main state-of-the-art results based on the RIP approach [3].

Theorem 2.2.1 (Noiseless Recovery) *Consider the system (1.2) with the unique solution \mathbf{s} , assume $\delta_{2k} < \sqrt{2} - 1$. Let $\hat{\mathbf{s}}$ be the solution to (1.4), then:*

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_1 \leq C_0 \|\mathbf{s} - \mathbf{s}_k\|_1$$

and

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_2 \leq C_0 \frac{1}{\sqrt{k}} \|\mathbf{s} - \mathbf{s}_k\|_1$$

where \mathbf{s}_k is a k -sparse approximation of \mathbf{s} and C_0 is a global constant.

Note that for the case the source is exactly k -sparse, this theorem states the recovery is exact. The next theorem states the condition for robustness towards noise.

Theorem 2.2.2 (Noiseless Recovery) *Consider the system $\mathbf{y} = A\mathbf{s} + \mathbf{n}$ such that $\|\mathbf{n}\|_2 < \epsilon$, assume $\delta_{2k} < \sqrt{2} - 1$. Let $\hat{\mathbf{s}}$ be the solution to (1.4), then:*

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_2 \leq C_0 \frac{1}{\sqrt{k}} \|\mathbf{s} - \mathbf{s}_k\|_1 + C_1 \epsilon$$

with the same constant C_0 as in the previous theorem and another global constant C_1 .

Proofs of these theorems are complicated and based on advanced real analysis mathematics. Interested readers may refer to [1–3] for details.

RIP condition on the sensing matrix is a standard approach in CS theory, but unfortunately its practical benefits is limited. Calculating RIP for a general matrix is a NP-hard problem and only has been done for special cases. Using random matrix theory, existence of such matrices have been proven for $m > O(k \log(\frac{n}{k}))$

for any desired $\delta_k \in (0, 1)$ but even in such cases building such matrices is an independent issue. Note these theorems require RIP condition but we will discuss that RIP condition is only a sufficient condition and is not a necessary condition for accurate l_1 -recovery. On the other hand, it is also not a complete concept to study CS.

An important quantity in designing the sensing matrix is mutual coherence.

Definition 4 *Mutual Coherence*

Let $A \in \mathbb{R}^{n \times m}$, the mutual coherence μ_A is defined by:

$$\mu_A = \max_{i \neq j} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\| \|\mathbf{a}_j\|}$$

where $\mathbf{a}_i, \mathbf{a}_j$ denote two distinct columns of A .

A small coherence implies of closeness of the sensing matrix to a normal matrix. If a matrix possesses a small mutual coherence, then it also satisfies the RIP condition. It means that coherence is a stronger condition. On the other hand the complexity of calculating the coherence is $O(n^2)$ and thus is tractable. According to Welch inequality [4]:

$$\mu_A \geq \sqrt{\frac{n}{m(n-m)}} \quad (2.17)$$

This implies for $n \gg m$, $\mu_A \geq \frac{1}{\sqrt{m}}$. Consequently if we want to design sensing matrices which satisfy RIP condition using mutual coherence, then $m > O(k^2)$ which is much greater than $m = O(k \log(\frac{n}{k}))$ bound for which existence of proper sensing matrices has been proven. But due to computational complexity issues, it is the only proper tool for this purpose.

Next section covers the new paradigm for compressive sensing [5–7]. This approach uses a completely different approach based on studying the nullspace of the sensing matrix using spherical section property.

2.3 Spherical Section Property

Analysis of compressive sensing based on RIP requires advanced mathematical tools, but this approach is not necessary to develop compressive sensing [5, 6]. Moreover, it is not a required condition for exact recovery.

Consider the problem of (1.2). The pair (A, \mathbf{y}) carries the information in CS framework. Consider an invertible matrix, B . It is trivial that the system $BA\mathbf{s} = B\mathbf{y}$ is equivalent to the system $A\mathbf{s} = \mathbf{y}$. Thus the pair $(BA, B\mathbf{y})$ carries the same information as (A, \mathbf{y}) . But the RIP of A and BA can be vastly different. For any CS problem one can choose B to make RIP of BA significantly bad, regardless of RIP of A [6]. RIP is a strong condition on sensing matrix and practical and experimental results confirm

it is not a necessary condition for main theorems of CS to hold. This motivates the derivation of CS in a more simple and general approach based on spherical section property (SSP) [5, 6]. Interestingly this approach is simpler and some of the main results of CS theorems in RIP context can be derived easier using spherical section property. Here we briefly describe CS theory in this context and follow the approach of [6] in proving the main theorems.

Definition 5 *Spherical Section Property (SSP)* Let $m, n \in \mathbb{N}$ such that $n > m$ and \mathbf{V} be an $n - m$ dimensional subspace of \mathbb{R}^n . This subspace is said to have spherical section property with constant Δ , if $\forall \mathbf{s} \in \mathbf{V}$:

$$\frac{\|\mathbf{s}\|_1}{\|\mathbf{s}\|_2} \geq \sqrt{\frac{m}{\Delta}}$$

Here, Δ is called the distortion of \mathbf{V} .

Note if we consider the nullspace of a sensing matrix as the subspace in this definition, for an invertible matrix $\Delta = 0$. Similar to RIP approach the following theorems are developed.

Theorem 2.3.1 (Noiseless Recovery) Suppose $\text{null}(\mathbf{A})$ has the Δ -spherical section property. Let $\hat{\mathbf{s}}$ be a nonzero vector such that: $\mathbf{A}\hat{\mathbf{s}} = \mathbf{y}$.

1. Provided that: $\|\hat{\mathbf{s}}\|_0 \leq \frac{m}{3\Delta}$, $\hat{\mathbf{s}}$ is the unique vector satisfying $\mathbf{A}\mathbf{s} = \mathbf{y}$ and $\|\mathbf{s}\|_0 \leq \frac{m}{3\Delta}$.
2. Provided that: $\|\hat{\mathbf{s}}\|_0 \leq \frac{m}{2\Delta} \leq \frac{n}{2}$, $\hat{\mathbf{s}}$ is the unique solution to the optimization problem (1.4).

Proof 1

1. First define the vector $\text{sign}(\mathbf{s}) = [\text{sign}(s_i)]$. According to the Cauchy-Schwarz inequality:

$$\begin{aligned} |\langle \text{sign}(\mathbf{s}), \mathbf{s} \rangle| &\leq \|\text{sign}(\mathbf{s})\|_2 \|\mathbf{s}\|_2 \rightarrow \sum_i |s_i| \leq \sqrt{|\text{supp}(\mathbf{s})|} \|\mathbf{s}\|_2 \rightarrow \|\mathbf{s}\|_1 \\ &\leq \sqrt{\|\mathbf{s}\|_0} \|\mathbf{s}\|_2 \end{aligned} \quad (2.18)$$

Now assume \mathbf{v} be a second solution which is more sparse compared to $\hat{\mathbf{s}}$ and $\|\mathbf{v}\|_0 = m_1$. Let $\mathbf{w} = \mathbf{v} - \hat{\mathbf{s}}$. Note, $\mathbf{w} \neq 0$ and $\mathbf{w} \in \text{Null}(\mathbf{A})$, then:

$$\begin{aligned} \|\mathbf{w}\|_0 &\leq \|\mathbf{v}\|_0 + \|\hat{\mathbf{s}}\|_0 \leq m_1 + \frac{m}{3\Delta} \xrightarrow{(2.18)} \frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2} \leq \sqrt{m_1 + \frac{m}{3\Delta}} \\ \sqrt{\frac{m}{\Delta}} &\leq \sqrt{m_1 + \frac{m}{3\Delta}} \rightarrow \frac{2m}{3\Delta} \leq m_1, \end{aligned} \quad (2.19)$$

this a contradiction and shows \mathbf{v} is not sparse enough and uniqueness of the solution results.

2. Again assume \mathbf{v} be a second solution to (1.4) such that $\|\mathbf{v}\|_1 \leq \|\hat{\mathbf{s}}\|_1$ and let $\mathbf{w} = \mathbf{v} - \hat{\mathbf{s}}$, $S = \text{supp}(\hat{\mathbf{s}})$, $\bar{S} = \{1, \dots, n\} - S$, and \mathbf{w}_S to be the projection of \mathbf{w} on S :

$$\|\mathbf{v}\|_1 = \|\mathbf{w} + \hat{\mathbf{s}}\|_1 = \|\mathbf{w}_S + \hat{\mathbf{s}}_S\|_1 + \|\mathbf{w}_{\bar{S}} + \hat{\mathbf{s}}_{\bar{S}}\|_1 = \|\mathbf{w}_S + \hat{\mathbf{s}}_S\|_1 + \|\mathbf{w}_{\bar{S}}\|_1 \geq \quad (2.20)$$

$$\|\hat{\mathbf{s}}_S\|_1 - \|\mathbf{w}_S\|_1 + \|\mathbf{w}_{\bar{S}}\|_1 = \|\hat{\mathbf{s}}\|_1 - \|\mathbf{w}_S\|_1 + \|\mathbf{w}_{\bar{S}}\|_1,$$

now since $\|\mathbf{v}\|_1 \leq \|\hat{\mathbf{s}}\|_1$, one concludes $\|\mathbf{w}_{\bar{S}}\|_1 \leq \|\mathbf{w}_S\|_1$.

Note $\mathbf{w} \in \text{null}(A)$, now we want to calculate maximum value of the ratio $\frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2}$. This problem is invariant under scaling of \mathbf{w} , thus we set $\|\mathbf{w}\|_2 = 1$ and also we can assume \mathbf{w} lays in the positive orthant (since the element signs would not change the norm value). We will have the following optimization problem:

$$\begin{aligned} \max \quad & w_1 + \dots + w_n \\ \text{s.t.} \quad & 0 \leq w_i, \\ & \sum_{i \in \bar{S}} w_i \leq \sum_{i \in S} w_i \end{aligned} \quad (2.21)$$

The second constraint comes from the inequality we derived before. This problem is a convex optimization instance, so we can exhibit the maximizer in closed form if we can exhibit the solution to the KKT condition [8]. Let

$$w_i = \begin{cases} a = \frac{\sqrt{\|\hat{\mathbf{s}}\|_0(n - \|\hat{\mathbf{s}}\|_0)/n}}{\|\hat{\mathbf{s}}\|_0}, & i \in S \\ b = \frac{\sqrt{\|\hat{\mathbf{s}}\|_0(n - \|\hat{\mathbf{s}}\|_0)/n}}{\|n - \hat{\mathbf{s}}\|_0}, & i \in \bar{S} \end{cases} \quad (2.22)$$

It is easy to check that this point lays in the feasible region. The KKT multipliers are the solutions to the system:

$$\begin{cases} \lambda_1 + 2\lambda_2 b = 1 \\ -\lambda_1 + 2\lambda_2 a = 1 \end{cases} \rightarrow \begin{cases} \lambda_1 = \frac{a-b}{a+b} \\ \lambda_2 = \frac{1}{a+b} \end{cases} \quad (2.23)$$

So both multipliers are positive if $\|\hat{\mathbf{s}}\|_0 \leq \|n - \hat{\mathbf{s}}\|_0$. Thus the objective value of (2.21) would be $\sqrt{\frac{\|\hat{\mathbf{s}}\|_0(n - \|\hat{\mathbf{s}}\|_0)}{n}}$ and consequently $\frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2} \leq \sqrt{\frac{\|\hat{\mathbf{s}}\|_0(n - \|\hat{\mathbf{s}}\|_0)}{n}}$. On the other hand $\mathbf{w} \in \text{null}(A)$, which concludes:

$$\sqrt{\frac{m}{\Delta}} \leq \sqrt{\frac{\|\hat{s}\|_0(n - \|\hat{s}\|_0)}{n}} \leq \sqrt{\|\hat{s}\|_0} \rightarrow \frac{m}{\Delta} \leq \|\hat{s}\|_0, \quad (2.24)$$

which contradicts the assumption and results the proof.

The second theorem considers stability towards noise.

Theorem 2.3.2 Noisy Recovery Suppose $\text{null}(A)$ has the Δ -spherical section property. Let \hat{s} be the minimizer of (1.4). Then for every $\bar{s} \in \mathbb{R}^n$ and $\forall k < \min(\frac{m}{16\Delta}, \frac{n}{4})$:

$$\|\hat{s} - \bar{s}\|_1 \leq 4\|\bar{s}_k - \bar{s}\|_1, \quad (2.25)$$

where s_k denotes the k -sparse approximation of s .

Proof 2 Let $w = \hat{s}_k - \bar{s}$, so $w \in \text{null}(A)$:

$$\begin{aligned} \|\hat{s}\|_1 &= \|\bar{s} + w\|_1 = \\ \|\bar{s}_S + w_S\|_1 + \|\bar{s}_{\bar{S}} + w_{\bar{S}}\|_1 &\geq \\ \|\bar{s}_S\|_1 - \|w_S\|_1 - \|\bar{s}_{\bar{S}}\|_1 + \|w_{\bar{S}}\|_1 &\geq \\ \|\bar{s}\|_1 - \|w_S\|_1 + \|w_{\bar{S}}\|_1 - 2\|\bar{s}_{\bar{S}}\|_1, \end{aligned} \quad (2.26)$$

Since \hat{s} is the minimizer of (1.4) we conclude:

$$\|w_{\bar{S}}\|_1 \leq \|w_S\|_1 + 2\|\bar{s}_{\bar{S}}\|_1. \quad (2.27)$$

Now define: $R = \frac{\|w\|_1}{\|\bar{s} - s_k\|_1}$. To obtain the result, it is enough to find an upper bound for R ($R \leq 4$). We substitute R in (2.27):

$$\begin{aligned} \|w_{\bar{S}}\|_1 \leq \|w_S\|_1 + 2\|w\|_1/R &\rightarrow \|w_{\bar{S}}\|_1 \leq \|w_S\|_1 + 2(\|w_S\|_1 + \|w_{\bar{S}}\|_1)/R \rightarrow \\ (1 - 2/R)\|w_{\bar{S}}\|_1 &\leq (1 + 2/R)\|w_S\|_1. \end{aligned} \quad (2.28)$$

Now note if $1 - 2/R \geq 0$, then $R \leq 2 \leq 4$ and the proof results, so let $1 - 2/R > 0$. Then from (2.28): $\|w_{\bar{S}}\|_1 \leq \frac{1+2/R}{1-2/R}\|w_S\|_1$. Assuming $\gamma = \frac{1+2/R}{1-2/R}$ ($\gamma \leq 3$) and in exactly the same approach as in the previous theorem one can conclude (for details refer to [6]):

$$\begin{aligned} \frac{\|w\|_1}{\|w\|_2} &\leq \gamma + \gamma \sqrt{\frac{k(n-k)}{k+9(n-k)}} \xrightarrow{w \in \text{null}(A)} \sqrt{\frac{m}{\Delta}} \leq \gamma + \gamma \sqrt{\frac{k(n-k)}{k+9(n-k)}} \\ &\xrightarrow{n-k \leq ((9(n-k)+k)/9)} \sqrt{\frac{m}{\Delta}} \leq (\gamma + 1)\sqrt{k} \xrightarrow{k \leq \frac{m}{16\Delta}} 3 \geq \gamma. \end{aligned} \quad (2.29)$$

On the other hand the assumption was $\gamma \leq 3$ and thus $\gamma = 3$. Consequently $R = 2$ which results in the desired bound on R and the result follows.

These two theorems establish CS theory but in SSP context and similarly state uniqueness and stability of l_1 -norm solution for a CS problem. The results are derived in a much simpler approach compared to RIP context [1, 2]. It is interesting to note that the main results which are derived in RIP approach can be rederived in SSP context. For instance the error bound in Theorem 2.3.2 has been derived in RIP context, too. Also, it has been shown the Gaussian random matrices have spherical section property and are good choice for sensing matrix [5]. Furthermore, as it will be discussed this approach is a better embed for considering cases when we have side information on the feasible region.

2.4 Reconstruction Methods

In this section a brief review on CS reconstruction methods is given. Nowadays one of the limitation of using CS is the low-speed of the reconstruction methods with high dimensional data. Improving the performance of reconstruction methods is an active research area.

2.4.1 Minimization of l_1 -norm

Historically l_1 -norm minimization is the main approach for CS reconstruction algorithms. Main CS theorems state robustness of the l_1 -norm minimization towards additive noise and also system noise. The importance of l_1 -norm is that, it is a continuous convex function, so convex optimization tools can be applied to the problem. The more important fact is that l_1 -norm minimization problem can be formulated as a linear programming problem. Let $A' = [A, -A]$, $\mathbf{s}' = [\mathbf{s}_1; \mathbf{s}_2]$, $\mathbf{s} = \mathbf{s}_1 - \mathbf{s}_2$:

$$\min[\mathbf{1}; \mathbf{1}]^T \mathbf{s}' \quad \text{s.t. } A' \mathbf{s}' = \mathbf{y}0, \mathbf{s}' \geq 0, \quad (2.30)$$

where $\mathbf{1}$ is an all-ones column vector and $(\cdot)^T$ denotes matrix transposition. Consequently well-known linear programming algorithms such as Simplex and Interior Point methods can be used with complexity of $O(n^3)$. One group of successful algorithms in this class is Basis Pursuit [9].

Although linear programming methods can find the solution in finite time but for many practical applications $O(n^3)$ is not a tractable time. Specially in image processing applications in which $n = O(10^5)$ for a typical image.

2.4.1.1 Thresholding Algorithms

Some iterative methods have been introduced to decrease the computational complexity of l_1 -norm minimization. In these methods an iterative sequence of vectors is produced, which converges to the solution through iterations. Although convergence to the exact solution is more time consuming compared to linear programming methods, these methods quickly converge to a very good approximate of the solution.

It can be shown that for a proper selection of λ , the optimization problem (1.4) is equivalent to the following unconstrained problem:

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \frac{1}{2} \|\mathbf{y} - A\mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1 \quad (2.31)$$

Since this problem is unconstrained one can use steepest descend or conjugate gradient approaches to derive an iterative relation. Although l_1 -norm is not a smooth function but concept of subderivative enables us to apply a similar procedure to steepest descend on (2.31) (more discussions is given in Chap.3). Upon choosing a proper initial value, the iterative relation will converge to the minimizer of (1.4). Several algorithms have been developed for this purpose [10, 11]. In the current note we work with image signals and thus we have used one of the-state-of-the-art iterative methods for reconstruction [12, 13].

The iterative formula for iterative hard thresholding (IHT) algorithm is as follows:

$$\mathbf{s}^{i+1} = \mathcal{G}(\mathbf{s}^i - A^T(A\mathbf{s}^i - \mathbf{y})), \quad (2.32)$$

where $\mathcal{G}(\cdot)$ is a thresholding function:

$$\mathcal{G}(x) = \begin{cases} 0 & |s_i| \leq \sqrt{\lambda} \\ s_i & |s_i| \geq \sqrt{\lambda} \end{cases}, \quad (2.33)$$

The main advantage is that each iteration only involves multiplication of vectors and A and A^T , followed by thresholding. So the sensing matrix can be defined only as an operator and it is not even required to store the sensing matrix. This is much simpler than linear programming. Note the threshold in this algorithm is constant in all iterations. A class of successful methods is the iterative shrinkage thresholding algorithms (ISTA) which improve IHT through using an adaptive thresholding function. The iterative step is as follows:

$$\mathbf{s}^{i+1} = \mathcal{H}_{\lambda\delta}(\mathbf{s}^i - \delta A^T(A\mathbf{s}^i - \mathbf{y})), \quad (2.34)$$

where δ is a parameter for step size and $\mathcal{H}(\cdot)$ is a soft shrinkage threshold function:

$$\mathcal{H}_{\lambda}(s_i) = (|s_i| - \lambda)_+ \text{sign}(x_i). \quad (2.35)$$

FISTA algorithm [12] further improves ISTA by involving the solutions of the two previous iterations in each step.

2.4.2 Greedy Algorithms

Greedy algorithms generally solve a problem in a number of steps (in CS problem, mainly the number of steps is equal to the sparsity level k). In each step the best selection (in CS problem, normally the best column of the sensing matrix) is done without considering the future steps. Consequently the result is not always the real solution but this approach provides acceptable results in compressive sensing reconstruction.

A simple algorithm of this class is Matching Pursuit. An equivalent representation for compressive sensing is:

$$\mathbf{y} = \sum_{i=1}^n \mathbf{a}_i s_i, \quad (2.36)$$

where \mathbf{a}_i is the i th column of A . If we have a k -sparse source, CS in this context can be interpreted as finding the k related columns of A and corresponding s_i 's. Matching Pursuit approximates the source in k step. In each step one column of A is revealed and then the corresponding s_i is revealed by solving a least square problem. In the first step the inner product of \mathbf{y} and all \mathbf{a}_i 's are calculated ($\langle \mathbf{y}, \mathbf{a}_i \rangle$). Then the column \mathbf{a}_j with the maximum absolute value of $\langle \mathbf{y}, \mathbf{a}_i \rangle$ is selected as an active column in (2.36) and $s_i = \langle \mathbf{y}, \mathbf{a}_j \rangle$. Thus the first term in (2.36) is known. Let this approximate of \mathbf{s} be $\mathbf{s}^{(1)}$. The next steps are done similarly, only in each step we update the value of \mathbf{y} as follows:

$$\mathbf{y}^{(i+1)} = \mathbf{y}^{(i)} - s_i \mathbf{a}_j. \quad (2.37)$$

The main disadvantage in this approach is that it is assumed that columns of A are orthogonal which is not the case for most sensing matrices. Orthogonal Matching Pursuit (OMP) [14] improves this method via updating the found s_i 's in each step. Since this approach uses similarity of \mathbf{a}_i 's and the residual vector of (2.37), mutual coherency of the sensing matrix plays an important role. Faster algorithms such as Compressive Sampling Matched Pursuit (CoSaMP) [15] improves the algorithm via a look on future steps. Overall, this class of reconstruction methods are fast but do not necessarily find the real solution.

2.4.3 Norm Approximation

This class approximate l_0 -norm via a differentiable function and then use methods such as steepest descend for minimization. For instance smoothed l_0 (SL0) algorithm [16] approximates the l_0 -norm as follows:

$$\|\mathbf{s}\|_0 \approx g(\mathbf{s}) = n - \sum_{i=0}^n f_\sigma(s_i), \tag{2.38}$$

where $f_\sigma(\cdot)$ is defined as:

$$f_\sigma(s) = e^{-\frac{s^2}{2\sigma^2}}, \tag{2.39}$$

and $\sigma \in \mathbb{R}^+$ is a small constant. The parameter σ determines the closeness to the l_1 -norm and smoothness of the approximation, as $\sigma \rightarrow 0$ then $g(\mathbf{s}) \rightarrow \|\mathbf{s}\|_0$. The function $g(\cdot)$ is continuous and differentiable and thus steepest descend methods can be applied directly to find the minimizer of $g(\cdot)$. For a proper selection of σ , it may be possible to find the global minimizer of (1.3). Experiments have shown that this method is faster than l_1 -norm minimization methods but again for large scale systems it is not applicable.

2.4.4 Message Passing Reconstruction Algorithms

Graphical Models is an active research area with a wide range of applications. Recently fast iterative methods based on graphical models have been used in convex optimization problems [17, 18]. The connection between belief propagation (BP) message passing algorithm and convex optimization inspired researchers to apply graphical models concepts to CS theory to find faster solvers.

In order to connect CS theory with graphical models, first we model CS problem as a probabilistic inference problem. Figure 2.1 [subplot (a)] provides a block diagram representation for (1.3). It is assumed that the sparse source is resulted

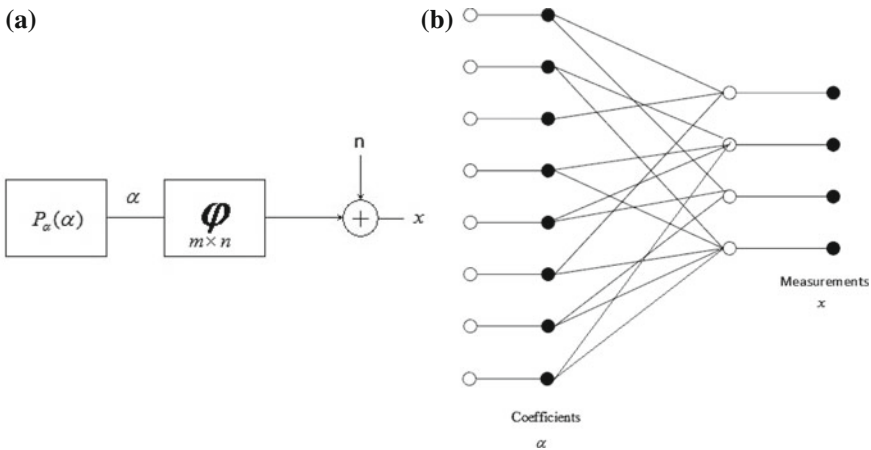


Fig. 2.1 (a) Probabilistic block diagram for CS and (b) corresponding factor graph

from sampling of a probability distribution $P_s(\mathbf{s})$. Sparse sources have been modeled with heavy-tailed distributions including Laplacian, Gaussian mixtures, generalized Gaussian, and Bernoulli Gaussian distributions in the literature [18]. The observation is resulted from the source via linear transformation, $A = \Phi\Psi$, followed by noise contamination. The goal is to estimate the source signal, either MAP or MMSE estimations, using the observed measurement, \mathbf{y} . In this framework the original CS problem can be considered as a probabilistic inference problem. Exact MAP estimate can be computed for the problem [18] but unfortunately the solution involves heavier computational load compared to l_1 -norm minimization methods. One idea is to use approximate inference algorithms such as BP to lessen the computational load. To do this end a graphical model must be assigned to the problem. The main idea for this purpose comes from error control coding area, where it is common to represent a parity check matrix by a bipartite graph. Analogously the block diagram in Fig. 2.1 [subplot (a)] can be represented by a bipartite factor graph as shown in Fig. 2.1 [subplot (b)]. There are two class of nodes in the factor graph: variable nodes (black) and constraint nodes (white). The edges connect variable nodes to constraint nodes. A constraint node models the dependencies that its neighboring variable nodes are subjected to. We have two types of constraint nodes; the first type imposes the probability distribution on source coefficients while the second type connects each coefficient node to a set of measurement variables that are used in computing that measurement. Having this factor graph, belief propagation can be employed to infer the probability distribution of the coefficients and consequently the MAP estimation for source signal.

In [18], the authors used belief propagation to infer the source signal. While their approach is interesting and the algorithm is much faster compared to general CS reconstruction algorithms, it poses a main limitation: to run BP, the authors assumed the sensing matrix to be sparse, which is not a realistic assumption in most CS applications. The reason for this assumption is that the implementation of BP in the general case is computationally intractable for dense graphs. Fortunately BP often admits acceptable solution for large, dense matrix when Gaussian approximation is used [19]. This property has led to generalization of approximate message passing algorithms for dense graphs. The key idea of generalized message passing algorithm (GMA) is to decompose the vector valued estimation problem into a sequence of scalar problems. This idea combined with the idea given in [18], has been used to generalize the compressive sensing algorithm via belief propagation for CS problems with dense sensing matrices. This class of algorithms are new compared to other classes and research is still going on to improve and generalize these algorithms to non-parametric cases, where we do not have prior information about the source distribution.

In this section a brief review on CS reconstruction algorithm was given. As stated in Chap. 1, one of the main limitations of applying CS to applications is at its reconstruction side. After about a decade of extensive research in this area, nowadays CS is well established and matured in terms of theory and analysis, but research is still going on to improve the current reconstruction algorithm in terms of computational

and implementational complexity. Simple algorithms which can be implemented cheaply via electronic devices is crucial for this research area.

References

1. E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theor.* **52**(2), 489–509 (2006)
2. Y. Tsaig, D.L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theor.* **52**, 1289–1306 (2006)
3. E.J. Candès, The restricted isometry property and its implications for compressed sensing. *C. R. Math.* **346**(9–10), 589–592 (2008)
4. T. Strohmer, R.W.J. Heath, Grassmannian frames with applications to coding and communication. *Appl. Comput. Harmonic Anal.* **14**(3), 257–275 (2003)
5. S.A. Vavasis, *Elementary proof of the spherical section property for random matrices* (University of Waterloo, Waterloo, Technical report, 2009)
6. Y. Zhang, On theory of compressive sensing via l_1 -minimization: Simple derivations and extensions. Technical Report TR08-11, Rice University, 2008
7. B. Kashin, V. Temlyakov, A remark on compressed sensing. *Math. Notes* **82**, 748–755 (2007)
8. A.W. Tucker, in *Proceedings of the second berkeley symposium on mathematical statistics and probability*, ed. by J. Neyman (University of California Press, Berkeley, 1951), pp. 481–492
9. S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit. *SIAM Rev.* **43**(1), 129–159 (2001)
10. E. van den Berg, M.P. Friedlander, Probing the pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* **31**(2), 890–912 (2008)
11. M.A.T. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Sign. Proces.* **1**(4), 586–597 (2007)
12. A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009)
13. C. Vonesch, M. Unser, A fast iterative thresholding algorithm for wavelet-regularized deconvolution, in *Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet XII*, vol 6701, (2007) pp 67010D–1–67010D-5
14. Y. C. Pati, R. Rezaifar, P. S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in *Proceedings of 27th Asilomar Conference on Signals Systems and Computers*, (1993) pp. 40–44
15. D. Needell, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Commun. ACM* **53**(12), 93–100 (2010)
16. G.H. Mohimani, M. Babaie-Zadeh, C. Jutten, A fast approach for overcomplete sparse decomposition based on smoothed l_0 norm. *IEEE Trans. Sig. Proces.* **57**(1), 289–301 (2009)
17. D. Needal, J. Tropp, Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harm. Anal.* **26**, 301–321 (2008)
18. D. Baron, S. Sarvotham, R.G. Baraniuk, Bayesian compressive sensing via belief propagation. *IEEE Trans. Sig. Proces.* **58**(1), 269–280 (2010)
19. M. Bayati, A. Montanari, The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theor.* **57**(2), 764–785 (2011)