“Compressive Sensing and Sampling” 学习笔记

黄冬勃

2015 年 4 月 1 日

Contents

Chapter 1

Study on "An Introduction To Compressive Sampling" [1]

CS theory asserts that one can recover certain signals and images from far fewer samples or measurements than traditional methods use. To make this possible, CS relies on two principles: *sparsity*, which pertains to the signals of interest, and *incoherence*, which pertains to the sensing modality.

1. ***Sparsity*** express the idea that the "information rate" of a continuous time signal may be much smaller than suggested by tis bandwidth, or that a discrete-time signal depends on a number of degrees of freedom which is comparably much smaller than its (finite) length. More precisely, CS exploits the fact that many natural signals are sparse or compressible in the sense that ey have concise representations when expressed in the proper basis Ψ

2. ***Incoherence*** extends the duality between time and frequency and expresses the idea that objects having a sparse representation in Ψ must be spread out in the domain in which they are acquired, just as a Dirac or a spike in the time domain is spread out in the frequency domain. Put differently, incoherence says that unlike the signal of interest, the sampling/sensing waveforms have an extremely dense representation in Ψ(密集表示？)

   > 非相关性拓展了时间与频率之间的二元性。并且表述了一个思想，即对象存在 Ψ 中有一个稀疏表达，该稀疏表达必须延伸至获取它们所在的域，就像时域上一个冲激响应函数 (Dirac) 或脉冲频域上的分布。非相关性说明了不像对于我们感兴趣的信号，采集/感知波形在 Ψ 中有着极其致密的表达。

The ***crucial observation*** is that one can design efficient sensing or sampling protocols that capture the useful information content embedded in a sparse signal and condense it into a small amount of data. ***These protocols*** are nonadaptive and simply require correlating the signal with a small number of fixed waveforms that are incoherent with the sparsifying basis.

What is the most remarkable about these sampling protocols is that they allow a sensor to very efficiently capture the information in a sparse signal without trying to comprehend that signal.

CS is a very simple and efficient signal acquisition protocol which samples–in a signal independent fashion–at a low rate and later uses computational power for reconstruction from what appears to be an incomplete set of measurements.

CS is a concrete protocol for sensing and compressing data simultaneously.

1.1 The Sensing Problem

The sensing mechanisms is discussed in which information about a signal $f(t)$ is obtained by linear functionals recording the values

$$y_k = \langle f, \varphi_k \rangle, \qquad k = 1, \cdots, m. \tag{1.1.1}$$

The object wish be acquired is simply correlated with waveforms $\varphi_k(t)$. This is a standard setup.

1. If the sensing waveforms are **Dirac delta functions(spikes, 冲激响应)**, then y is a vector of sampled values of f in the time or space domain.

2. If the sensing waveforms are **indicator functions**($1_A(x) = \begin{cases} 1 & 若\ x \in A, \\ 0 & 若\ x \notin A. \end{cases}$) of pixels, then y is the image data typically collected by sensors in a digital camera.

3. If the sensing waveforms are sinusoids, then y is a vector of Fourier coefficients;

this is the sensing modality used in *magnetic resonance imaging* (MRI). The *undersampled*[1] situations, in which the number m of available measurements is much smaller than dimension n of the signal f, are extremely common for a variety of reasons:

1. number of sensors may be limited;

2. the measurements may be extremely expensive as in certain imaging processes via *neutron scattering(中子散射)*.

This state of affairs looks rather daunting, as one would need to solve an **underdetermined linear system of equations(欠定线性系统方程组)**.

Letting A denote the $m \times n$ sensing matrix with the vectors $\varphi_1^*, \cdots, \varphi_m^*$ as rows (a_* is the **complex transpose**(**复转置, complex conjugate transpose–复共轭转置**) of a), the process of recovering $f \in \mathbb{R}^n$ from $y = Af \in \mathbb{R}^m$ is *ill-posed(不适定的)* in general when $m < n$: there are infinitely many candidate signals $\tilde{f}$ for which $A\tilde{f} = y$.

By relying on naturally existed realistic models of objects f, a way out could perhaps be imagined.

Shannon theory tells us that, if $f(t)$ actually has very low bandwidth, then a small number of (uniform) samples will suffice for recovery.

1.2 Incoherence and The Sensing of Sparse Signals

1.2.1 Sparsity

Many natural signals have concise representations when expressed in a convenient basis. Although nearly all the image pixels have nonzero values, the wavelet coefficients offer a concise summary: *most coefficients are small, and the relatively few large coefficients capture most of the information.*

Mathematically, a vector $f \in \mathbb{R}^n$ is expanded in an **orthonormal basis (标准正交基 such as a wavelet basis)** $\Psi = [\psi_1 \psi_2 \cdots \psi_n]$ as follows:

$$f(t) = \sum_{i=1}^{n} x_i \psi_i(t), \qquad t = 1, \cdots, N. \tag{1.2.1}$$

1. where x is the *coefficient sequence* of f, $x_i = \langle f, \psi_i \rangle$.

2. It will be convenient to express f as Ψx (where Ψ is the $n \times n$ matrix with $\psi_1, \cdots, \psi_n$ as columns).

The implication of sparsity:**when a signal has a sparse expansion, one can discard the small coefficients without much perceptual loss**.

Formally, consider $f_S(t)$ obtained by keeping only the terms corresponding to the S largest values of (x_i) in the expansion eq. (1.2.1). By definition, $f_S := \Psi x_S$, x_S is the vector or coefficients (x_i) with all but teh largest S set to zero. *S-sparse* denotes that objects with at most S nonzero entries.

Since Ψ is an orthonormal basis (or “orthobasis”), we have $\| f - f_S \|_{\ell 2} = \| x - x_S \|_{\ell 2}$, and if x is sparse or compressible in the sense that the sorted **magnitudes of the** (x_i) **decay quickly**, then x is well approximated by x_S and, therefore, the error $\| f - f_S \|_{\ell 2}$ is small.

[1]欠采样

A simple method for data compression would be to compute x from f and then (adaptively) encode the locations and values of the S significant coefficients. Such a process requires knowledge of all the n coefficients x, as the locations of the significant pieces of information may not be known in advance (they are signal dependent). More generally, sparsity is a fundamental modeling tool which permits efficient fundamental signal processing. Sparsity *has significant bearings on(与……紧密相关)* the acquisition process itself. **Sparsity determines how efficiently one can acquire signals nonadaptively**.

1.2.2 Incoherent Sampling

Given a pair (Φ, Ψ) of orthobases of $\mathbb{R}^n$.

- **Φ is used for sensing the object f as in eq. (1.1.1)–*sensing basis***.
- **Ψ is used to represent f–*representation basis***

The restriction to pairs of orthobases is **not essential** and will **merely simplify our treatment**.

Definition 1.2.1. *The coherence between the sensing basis Φ and the representation basis Ψ is*

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \le k, j \le n} |\langle \varphi_k, \psi_j \rangle| . \tag{1.2.2}$$

In plain English, the coherence measures the largest correlation between any two elements of Φ and Ψ, see also [2]. If Φ and Ψ contain correlated elements, the coherence is large. Otherwise, it is small. As for how large and how small, it follows from linear algebra that $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$.

Compressive sampling is mainly concerned with *low coherence pairs*, and examples of such pairs are given. In first example, Φ is the *canonical or spike basis(标准或者冲击响应基)* $\varphi(t) = \delta(t-k)$ and Ψ *is the Fourier basis*, $\Psi_j(t) = n^{-1/2} e^{i2\pi jt/n}$. Φ is the sensing matrix, this corresponds to the classical sampling scheme in time or space. The time-frequency pair obeys $\mu(\Phi, \Psi) = 1$, therefore, we have ***maximal incoherence***. Further, spikes and sinusoids are maximally incoherent *not just in one dimension but in any dimension*, (in two dimensions, three dimensions, etc.)

In second example, wavelets bases for Ψ and noiselets [3] are taken. The coherence between noiselets and Haar wavelets is $\sqrt{2}$ and that between *Daubechies D4 and D8 wavelets* is, respectively, about 2.2 and 29 across a wide range of sample sizes n. This extends to higher dimensions as well. (Noiselets are also maximally incoherent with spikes and incoherent with the Fourier basis.)

The interest in **noiselets** comes from the fact that:

1. they are incoherent with systems providing sparse representations of image data and other typers of data;
2. they come with very fast algorithms;
3. the noiselet transform runs in $O(n)$ time, and just like the Fourier transform, the noiselet matrix does not need to be sotred to be applied to a vector.

This is ***of crucial practical importance for numerical efficient CS implementations***.

Finally, random matrices are largely incoherent with any fixed basis Ψ. Select an orthobasis Φ uniformly at random, which can be done by

orthonormalizing n vectors sampled independently and uniformly on the unit sphere.

$$\Longrightarrow$$

With high probability, the coherence between Φ and Ψ is about $\sqrt{2 \log n}$

By extension, random waveforms $(\varphi_k(t))$ with independent identically distributed (i.i.d.)[2] entries, e.g., Gaussian or ± 1 binary entries, will also exhibit a very low coherence with any fixed representation Ψ.

[2] 独立同分布 (independent identically distributed,IID)

Note the rather strange implication here; if sensing with incoherent systems is good, then efficient mechanisms ought to acquire correlations with random waveforms, e.g., white noise! 应注意一个相当奇怪的涵义，即如果非相关系统的感知效果较好，那么有效的机制应该是获取随机波形的相关性，例如，白噪声。

1.2.3 Undersampling and Sparse Signal Recovery

It is only to get to observe a subset of the n *coefficients of* f and collect the data

$$y_k = \langle f, \varphi_k \rangle, \qquad k \in M, \tag{1.2.3}$$

where $M \subset \{1, \cdots, n\}$ is a subset of cardinality $m < n$.

So in this paper [1], the author decided to recover the signal by $\ell_1 - norm$ minimization; the proposed reconstruction f^* is given by $f^* = \Psi x^*$ is the solution to the *convex optimization*[3] program ($\|x\|_{\ell_1} := \sum_i |x_i|$)

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \qquad \text{subject to} \qquad y_k = \langle \varphi_k, \Psi \tilde{x} \rangle, \qquad \forall k \in M. \tag{1.2.4}$$

That is, among all objects $\tilde{f} = \Psi \tilde{x}$ consist with the data, we pick that whose coefficient sequence has minimal ℓ_1 norm.[4]

A leading early application of ℓ_1 norm as a *sparsity-promoting*[5] function was reflection seismology[6], in which a sparse reflection function(indicating meaningful changes between subsurface layers[7]) was sought from bandlimited data [4, 5]. However, ℓ_1-minimization is not the only way to recover sparse solutions; other methods, *such as greedy algorithms* [6], have also been proposed.

Theorem 1.2.1. *[7] Fix $f \in \mathbb{R}^n$ and suppose that coefficient sequence x of f in the basis Ψ is* S-sparse. *Select m measurements in the Φ domain uniformly at random. Then if*

$$m \geqslant C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot \log n \tag{1.2.5}$$

for some positive constant C, the solution to eq. (1.2.4) *is exact with overwhelming probability. (It is shown that the probability of success exceeds $1 - \delta$ if $m \geqslant C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot \log(n/\delta)$. In addition, the result is only guaranteed for nearly all sign sequences x with a fixed support, see [7] for details.)*

Three comments are made:

1) The role of the coherence is completely transparent: *the smaller the coherence, the fewer samples are needed*, so in the previous section the emphasis was on low coherence systems.

2) One suffers no information loss by measuring just about any set of m coefficients which may be far less than the signal size apparently demands. If $\mu(\Phi, \Psi)$(eq. (1.2.2)) is equal or close to one, then on the order of $S \log n$ samples suffice instead of n.

3) The signal f can be exactly recovered from our condensed data set by minimizing a *convex functional*[8] which does not assume any knowledge about the number of nonzero coordinates of x, their locations, or their amplitudes which we assume are all completely unknown a priori. We just run the algorithm if the signal happens to be sufficiently sparse, exact recover occurs.

The theorem indeed suggests a very concrete acquisition protocol: *sample nonadaptively in an incoherent domain (在非相干域进行非自适应采样)* and *invoke linear programming after acquisition step (采集步骤过后调用线性规划（编程?）)*. Then the signal in a compressed form would be essentially acquired. All that is needed is a decoder to "decompress" this data; **this is the role of ℓ_1 minimization**.

[3] 凸优化

[4] As is well known, minimizing ℓ_1 subject to linear equality constraints can easily be recast as a linear program making available a host of ever more efficient solution algorithms.

[5] 稀疏推进? 促进?

[6] 地震学

[7] 有较大意义的地下层变化指示

[8] 凸泛函

This random incoherence sampling theorem extends an earlier result about the sampling of spectrally sparse signals [8], which showed that randomness

1) *can be a very effective sensing mechanism and*

2) *is amenable to rigorous proof,*

and thus perhaps **triggered the many CS developments** we have witnessed and continue to witness today.

Suppose that we are interested in sampling ultra-wideband but spectrally sparse signals of the form $f(t) = \sum_{j=0}^{n-1} x_j e^{i2\pi jt/n}$ $t = 0, \cdots, n-1$, where n is very large but where the number of nonzero components x_j is less than or equal to S(which should be considered comparably small). Because the active set is not necessarily a subset of consecutive integers, the *Nyquist/Shannon theory is mostly unhelpful* (since one cannot restrict the bandwidth a priori, one may be led to believe that all n time samples are needed). In this special instance, theorem 1.2.1 claims that one can reconstruct a signal *with arbitrary and unknown frequency* support of size S from *on the order of* $S \log n$[9] time samples, see [8]. What is more, these samples do not have to be carefully chosen; almost any sample set of this size will work. For other types of theoretical results in this direction using completely different ideas see [9–11].

The point of the role played by probability is that to get useful and powerful results, one needs to resort to a probabilistic statement since one can not hope for comparable results holding for all measurement sets of size m. The reason: there are special sparse signals that vanish nearly everywhere in the Φ domain.

> In other words, one can find sparse signals f and very large subsets of size almost n (e.g., $n - S$) for which $y_k = \langle f, \varphi_k \rangle = 0$[10] for all $k \in M$.

Dirac comb[11] examples are discussed in [12] and [8].

- Given such subsets, one would get to see a stream of zeros and no algorithm whatsoever would of course be able reconstruct the signal.

- The theorem 1.2.1 guarantees that the fraction of sets for which exact recovery does not occur is truly negligible (a large negative power of n).

Thus, we only have to tolerate a probability of failure that is extremely small. For practical purposes, the probability of failure is zero provided that the sampling size is sufficiently large.

Through the study of special sparse signals discussed above, it is shown that one needs at least on the order of $\mu^2 \cdot S \cdot \log n$ samples as well.[12] Take $2S$ consecutive time points, see section 1.5, [9,10]. With fewer samples than $S \log n$, no matter how intractable, the probability that information may be lost is just too high and reconstruction by any method.

In summary, ***when the coherence is one, we do not need more than*** $S \log n$ ***samples but we cannot do with fewer either***.

The example in Figure 1(c) [1], the considered sparse image has only 25000 nonzero wavelet coefficients, with 96000 incoherent measurements to recovery information, the eq. (1.2.4) is solved (see [7] for the particulars of these measurements).

> The minimum-ℓ_1 recovery is perfect; that is, $f^* = f$.

There is *de facto*[13] a known four-to-one practical rule which says that for exact recovery, one needs about four incoherent samples per unknown nonzero term.

[9] 近似 $S \log n$
[10] 稀疏后的采集数据，为 0 表明该位置稀疏采样值为 0
[11] 狄拉克 (冲激响应) 梳状函数
[12] There exist subsets of cardinality $2s$ in the time domain which can reconstruct any s-sparse signal in the frequency domain.
[13] 事实上

1.3 Robust Compressive Sampling

In order to powerfully recover sparse signals from just a few measurements, CS needs to be able to deal with both *nearly sparse signals* and with *noise*.

This section examines two issues simultaneously:

1. First, general objects of interest are not exactly sparse but approximately sparse. The issue here is whether or not it is possible to obtain accurate reconstructions of such objects from highly undersampled measurements.
2. Second, in any real application measured data will invariably be corrupted by at least a small amount of noise as sensing devices do not have infinite precision.

It will ease the exposition to consider the abstract problem of recovering a vector $x \in \mathbb{R}^n$ from data

$$y = Ax + z, \tag{1.3.1}$$

where A IS an $m \times n$ “sensing matrix” giving us information about x, and z is a stochastic or deterministic unknown error term.

$$f = \Psi x$$

$$\&\qquad y = R\Phi f$$

(where R is the $m \times n$ matrix extracting the sampled coordinates in M)

$$\Longrightarrow$$

$$y = Ax,\ where\ A = R\Phi\Psi.$$

Hence, one can work with the abstract model eq. (1.3.1) bearing in mind that x may be the coefficient sequence of the object in a proper basis.

1.3.1 Restricted Isometries (sections 2.3.2 and 5.2.4)

In this section, **a key notion that has proved to be very useful to study the general robustness of CS** is introduced; the so-called ***restricted isometry property (RIP)***[14] [13].

Definition 1.3.1. *For each integer $S = 1, 2, \cdots$, define the isometry constant δ_S of a matrix A as the smallest number such that*

$$(1-\delta_S)\|x\|_{\ell_2}^2 \le \|Ax\|_{\ell_2}^2 \le (1+\delta_S)\|x\|_{\ell_2}^2 \tag{1.3.2}$$

holds for all S-sparse vectors x.

It can be loosely said that a matrix A obeys the RIP of order S if δ_S is not too close to one.

Note,

This property section 1.3.1 holds,

$$\Longrightarrow$$

A approximately preserves the Euclidean length of S-sparse signals,

$$\Longrightarrow$$

implies that S-sparse vecotrs cannot be in the null space of A.

This is useful as otherwise there would be no hope of reconstructing these vectors. An equivalent description of the RIP is to say that all subsets of S columns taken from A are in fact *nearly orthogonal*[15].

To see connection between the RIP and CS, imagine S-sparse signals with A wish to be acquired. Suppose that δ_{2S} is sufficiently less than one.

[14] 约束等距性，论文 [13] 看下

[15] the columns of A cannot be exactly orthogonal since we have more columns than rows

This implies that all pairwise distances between S-sparse signals must be well preserved in the measurement space. That is

$$(1-\delta_{2S})\|x_1-x_2\|_{\ell_2}^2 \leq \|Ax_1-Ax_2\|_{\ell_2}^2 \leq (1+\delta_{2S})\|x_1-x_2\|_{\ell_2}^2$$

holds for all S-sparse vectors $x_1,\ x_2$.

1.3.2 General Signal Recovery From Undersampled Data

If the RIP (section 1.3.1) holds, then the following linear program gives an accurate reconstruction:

$$\min_{\tilde{x}\in\mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \qquad \text{subject to} \qquad A\tilde{x} = y(=Ax). \tag{1.3.3}$$

Theorem 1.3.1. *[14] Assume that $\delta_{2S} < \sqrt{2}-1$. Then the solution x^* to eq. (1.3.3) obeys*

$$\begin{aligned} \|x^*-x\|_{\ell_2} &\leq C_0 \cdot \|x-x_S\|_{\ell_1}/\sqrt{S} \quad \textit{and} \\ \|x^*-x\|_{\ell_1} &\leq C_0 \cdot \|x-x_S\|_{\ell_1} \end{aligned} \tag{1.3.4}$$

for some constant C_0, where X_S is the vector x with all but the largest S components set to 0. [16].

The conclusions of theorem 1.3.1 are stronger than those of ??. If x is **not** S-sparse, then $x = x_S$ and, thus, the recovery is exact. If x is not S-sparse, then theorem 1.3.1 asserts that the quality of the recovered signal is as good as if one knew ahead of time the location of the S largest values of x and decided to measure those directly. The reconstruction is nearly as good as that provided by an oracle which, with full and perfect knowledge about x, extracts the S most significant pieces of information for us.

What is missing at this point is the relationship between S[17] obeying the hypothesis and m the number of measurements or rows of the matrix. To derive powerful results, we would like to find **matrices obeying the RIP with values of S close to m**.

1.3.3 Robust Signal Recovery From Noisy Data

It is given noisy data as in eq. (1.3.1) and use ℓ_1 minimization with relaxed constraints for reconstruction:

$$\min \|\tilde{x}\|_{\ell_1} \qquad \text{subject to} \qquad \|A\tilde{x}-y\|_{\ell_2} \leq \epsilon \tag{1.3.5}$$

where ϵ bounds the amount of noise in the data. (One could also consider recovery programs such as the *Dantzig selector* [17] or a combinatorial optimization program proposed by Haupt and Nowak [18]; both algorithms have provable results ***in the case where the noise is Gaussian with bounded variance***.) Problem eq. (1.3.5) is often called the **LASSO** after [19]; see also [20]. To the best of the author's knowledge, it was first proposed in [5]. This is again a convex problem (a second-order cone program) and can be solved efficiently.

Theorem 1.3.2. *[14] Assume that $\delta_{2S} < \sqrt{2}-1$. Then the solution x^* to eq. (1.3.5) obeys*

$$\|x^*-x\|_{\ell_2} \leq C_0 \cdot \|x-x_S\|_{\ell_1}/\sqrt{S} + C_1 \cdot \epsilon \tag{1.3.6}$$

for some constants C_0 and C_1. (Unpublished as stated and is a variation on the result found in [14]).

This can hardly be simpler. The reconstruction error is bounded by the sum of two terms:

1) The first is the error which would occur if one had noiseless data.

2) The second is just proportional to the noise level.

Further, C_0 and C_1 are typically small. With $\delta_{2S} = 1/4$ for example, $C_0 \leq 5.5$ and $C_1 \leq 6$.

This last result establishes CS as a practical and robust sensing mechanism. It works with all kinds of not necessarily sparse signals, and it handles noise gracefully. What remains to be done is to design efficient sensing matrices obeying the RIP. This is *the subject of*[18] the next sectionsection 1.4.

[16] As stated, this result is due to the first author [15] and yet unpublished, see also [14] and [16]

[17] the number of components one can effectively recover

[18] 以……为主题。on the subject of–论及；以……为课题

1.4 Random Sensing

Returning to the RIP, we would like to find sensing matrices with the property that column vectors taken from arbitrary subsets are nearly orthogonal. The larger these subsets, the better.

This is where randomness re-enters the pictures. Consider the following sensing matrices:

i) form A by sampling n column vectors uniformly at random on the unit sphere of $\mathbb{R}^m$;

ii) form A by sampling i.i.d.section 1.2.2 entries from the normal distribution with mean 0 and variance $1/m$;

iii) form A by sampling a random projection P as in "Incoherent Sampling" and normalize: $A = \sqrt{n/m}P$;

iv) form A by sampling i.i.d.section 1.2.2 entries from a symmetric Bernoulli distribution ($P(A_{i,j} = \pm 1/\sqrt{m}) = 1/2$) or other sub-Gaussian distribution.

With overwhelming probability, all these matrices obey the RIPsection 1.3.1 (i.e. the condition of our theoremtheorem 1.3.2) provided that

$$m \geq C \cdot S \log(n/S), \tag{1.4.1}$$

where C is some constant depending on each instance. The claims for i)-iii) use fairly standard results in probability theory; arguments in iv) are moire subtle[19] [21, 22]. In all these examples, the probability of sampling a matrix not obeying the RIP when eq. (1.4.1) holds is exponentially small in m. Interestingly, there are no measurement matrices and no reconstruction algorithm whatsoever which can give the conclusions of theorem 1.3.1 with substantially fewer samples than the left-hand side of eq. (1.4.1) [8, 23]. In that sense, using randomized matrices together with ℓ_1 minimization is a near-optimal sensing strategy.

One can also establish the RIP for pairs of orthobases as in "Incoherence and the Sensing of Sparse Signals". With $A = R\Phi\Psi$ where *R extracts m coordinates uniformly at random*, it is sufficient to have

$$m \geq C \cdot S(\log n)^4, \tag{1.4.2}$$

for the property to hold with large probability [8, 24]. If one wants a probability of failure no larger than $O(n^{-\beta})$ for some $\beta > 0$, then the best known exponent in eq. (1.4.2) is **five** instead of **four** (it is believed that eq. (1.4.2) holds with just $\log n$). ***This proves that one can stably and accurately reconstruct nearly sparse signals from dramatically undersampled data in an incoherent domain.***

Finally, the RIP can also hold for sensing matrices $A = \Phi\Psi$, where Ψ is an arbitrary orthobasis and Φ is an $m \times n$ measurement matrix drawn randomly from a suitable distribution.

If one fixes Ψ and populates Φ as in i)–iv), then with overwhelming probability, the matrix $A = \Phi\Psi$ obeys the RIP provided that eq. (1.4.1) is satisfied, where again C is some constant depending on each instance.

These random measurement matrices Φ are in a sense *universal* [21]; the sparsity basis need not even be known when designing the measurement system!

1.5 What is Compressive Sampling

At the compression stage massive amount of data are collected only to be–in large part–discarded.

Extremely wasteful process of massive data acquisition followed by compression:

1. one acquires a high-resolution pixel array f,
2. computes the complete set of transform coefficients,
3. encode the largest coefficients and discard all the others,
4. essentially ending up with f_S.

[19] [21] 看一下

CS operates very differently, and performs as "if it were possible to directly acquire just the important information about the object of interest".

Take about $O(S \log(n/S))$ random projections as in "Random Sensing",

$$\Longrightarrow$$

enough information to reconstruct the signal with accuracy at least as good as that provided by f_S,

$$\Longrightarrow$$

the best S-term approximation–the best compressed representation–of the object.

Or,

> CS measurement protocols essentially translate analog data into an already compressed digital form so that one can—-at least in principle-—obtain **super-resolved signals**[20] from just a few sensors.

There are some superficial similarities between CS and ideas in coding theory and more precisely with the theory and practice of Reed-Solomon (RS) codes [25]. **One can adapt ideas from coding theory to establish the following:**

> one can uniquely reconstruct any S-sparse signal from the data of its first $2S$ *Fourier coefficients*, $y_k = \sum_{t=0}^{n-1} x_t e^{-i2\pi kt/n}$, $k = 0, 1, 2, \cdots, 2S-1$, or from any set of $2S$ consecutive frequencies for that matter (the computational cost for the recovery is essentially that of solving an S × S Toeplitz system and of taking an n-point fast Fourier transform).

But this technique is not suitable to sense compressible signals. There are two main reasons:

1. First, the problem is that RS decoding is an algebraic technique, which cannot deal with nonsparse signals (the decoding finds the support by rooting a polynomial);
2. second, the problem of finding the support of a signal — even when the signal is exactly sparse — from its first 2S Fourier coefficients is extraordinarily ill posed (the problem is the same as that of extrapolating a high degree polynomial from a small number of highly clustered values).

Tiny perturbations of these coefficients will give completely different answers so that with finite precision data, reliable estimation of the support is practically impossible.

Whereas purely algebraic methods

1. **ignore the conditioning of information operators**,
2. **having well-conditioned matrices**,

which are crucial for accurate estimation, is a **central concern in CS as evidenced by the role played by the RIP**.

1.6 Applications

Data Compression The sparse basis Ψ may be unknown at the encoder or impractical to implement for data compression. As what is discussed in "Random Sensing". However, because a randomly designed Φ need not be with regards to the structure of Ψ, it can be considered a universal encoding strategy. (The knowledge and ability to implement Ψ are required only for the decoding or recovery of f.) This universality may be particularly helpful for distributed source coding in multi-signal settings such as sensor networks [26].

Channel Coding As explained in [13], CS principles (sparsity, randomness, and convex optimization) can be turned around and applied to design *fast error correcting codes over the reals to protect from errors* during transmission.

[20] 超分辨信号

Inverse Problems In still other situations, the only way to acquire f may be to use a measurement system Φ of a certain modality. However, assuming a sparse basis Ψ exists for f that is also incoherent with Φ, then efficient sensing will be possible. One such application involves MR angiography [8] and other types of MR setups [27], where Φ records a subset of the Fourier transform, and the desired image f is sparse in the time or wavelet domains.

Data Acquisition Here, it could be helpful to design physical sampling devices that directly record discrete, low-rate incoherent measurements of the incident analog signal.

A CS camera that collects incoherent measurements using a digital micromirror array (and requires just one photosensitive element instead of millions) could significantly expand these capabilities. (See [28] and an article by Duarte et al. in this issue.)

Two specific architectures for A/I [21] was proposed, in which a discrete, low-rate sequence of incoherent measurements can be acquired from a high-bandwidth analog signal. ***Each measurement y_k can be interpreted as the inner product $\langle f, \varphi_k \rangle$ of the incident analog signal f against an analog measurement waveform φ_k.*** As in the discrete CS framework, our preliminary results suggest that analog signals obeying a sparse or compressible model (in some analog dictionary Ψ) can be captured efficiently using these devices at a rate proportional to their information level instead of their Nyquist rate. There are challenges one must address when applying the discrete CS methodology to the recovery of sparse analog signals. The two architectures are as follows:

Nonuniform Sampler (NUS) The first architecture simply digitizes the signal at randomly or pseudo-randomly sampled time points. That is,

$$y_k = f(t_k) = \langle f, \delta_{t_k} \rangle .$$

In effect, these time points are obtained by jittering nominal (low-rate) sample points located on a regular lattice.

> Due to the incoherence between spikes and sines, this architecture can be used to sample signals having sparse frequency spectra far below their Nyquist rate.

There are tremendous benefits associated with a reduced sampling rate, as this provides added circuit settling time and has the effect of reducing the noise level.

Random Modulation Preintegration (RMPI) The second architecture is *applicable to a wider variety of sparsity domains*, most notably those signals having a sparse signature in the time-frequency plane. *Whereas* it may not be possible to digitize an analog signal at a very high rate, it may be quite possible to change its polarity at a high rate.

> The idea of the RMPI architecture is then to multiply the signal by a pseudo-random sequence of $\pm 1s$, integrate the product over time windows, and digitize the integral at the end of each time interval.

This is a parallel architecture and one has several of these random multiplier-integrator pairs running in parallel using distinct sign sequences. In effect, the RMPI architecture correlates the signal with a *bank of sequences of ± 1, one of the random CS measurement processes known to the universal, and therefore the RMPI measurements will be incoherent with any fixed time-frequency dictionary* such as the Gabor dictionary described below.

For each of the above architectures, the author have confirmed numerically (and in some cases physically) that the system is robust to *circuit nonidealities* such as *thermal noise*, *clock timing errors*, *interference*, and *amplifier nonlinearities.*

The application of A/I architectures to realistic acquisition scenarios will require continued development of CS algorithms and theory. An final discrete example is concluded:

> Take f to be a one-dimensional signal of length $n = 512$ that contains two modulated pulses. **From this signal**, we collect $m = 30$ measurements using an $m \times n$ measurement matrix Φ populated with i.i.d. Bernoulli

[21] analog-to-information

±1 entries. This is an unreasonably small amount of data corresponding to an undersampling factor of over 17. **For reconstruction** we consider a Gabor dictionary Ψ that consists of a variety of sine waves time limited by Gaussian windows, with different locations and scales. Overall the dictionary is approximately $43\times$ overcomplete and does not contain the two pulses that comprise of f.

There is more information on these directions in [29][22].。

[22]看一下该文章 [29]

Chapter 2

"Compressive Sampling" Notes

2.1 Introduction

Nyquist/Shannon sampling theory:

> the number of samples needed to reconstruct a signal without error is dictated by tis bandwidth – the length of the shortest interval which contains the support of the specturm of the signal under study.

An enchanting aspect of compressive sampling it that it has significant interactions and bearings on some fields in the applied sciences and engineering such as statistics, information theory, coding theory, theoretical computer science, and others as well.

- Sparsity leads to efficient estimations; *for example*, the quality of estimation by thresholding or shrinkage algorithms depends on the sparsity of the signal we wish to estimate.
- Sparsity leads to efficient compression; *for example*, the precision of a transform coder depends on the sparsity of the signal we wish to encode [30].
- Sparsity leads to dimensionality reduction and efficient modeling.

The *sparsity* has bearings on[1] *the data acquisition process itself, and leads to efficient data acquisition protocols.*

compressive sampling suggests ways to economically translate analog data into already compressed digital form [23, 31].

2.2 Undersampled Measurements

Consider the general problem of reconstructing a vector $x \in \mathbb{R}^N$ from linear measurements y about x of the form (same as eq. (1.1.1))

$$y_k = \langle x, \varphi_k \rangle , k = 1, 2, \cdots , K, \qquad or \qquad y = \Phi x \tag{2.2.1}$$

That is, information about the unknown signal can be acquired by sensing x against K vectors $\varphi_k \in \mathbb{R}^N$.

2.2.1 A nonlinear sampling theorem

Suppose here that one collects an incomplete set of frequency samples of a discrete signal x of length N. (To ease the exposition, we consider a model problem in one dimension. The theory extends easily to higher dimensions. For instance, we could be equally interested in the reconstruction of 2- or 3-dimensional objects from undersampled Fourier data.) The goal is to reconstruct the full signal f given only K samples in the Fourier domain

$$y_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_t e^{-i2\pi\omega_k t/N}, \tag{2.2.2}$$

[1] 与……有关；对……有影响

where the 'visible' frequencies ω_k are a subset Ω (of size K) of the set of all frequencies $\{0, \cdots, N-1\}$. In the language of the general problem eq. (2.2.1), the sensing matrix Φ is obtained by sampling K rows of the N by N discrete Fourier transform matrix.

A vector is S-sparse if its support $\{i : x_i \neq 0\}$ is of cardinality less or equal to S. Then [8] showed that one could almost always recover the signal x exactly by solving the convex program[2] ($\|\tilde{x}\|_{\ell_1} := \sum_{i=1}^{N} |\tilde{x}_i|$)

$$(P_1) \qquad \min_{\tilde{x}\in\mathbb{R}^N} \|\tilde{x}\|_{\ell_1} \qquad \text{subject to} \qquad \Phi\tilde{x} = y. \tag{2.2.3}$$

Theorem 2.2.1 ([8]). *Assume that x is S-sparse and that we are given K Fourier coefficients with frequencies selected uniformly at random. Suppose that the number of observations obeys*

$$K \geq C \cdot S \cdot \log N. \tag{2.2.4}$$

Then minimizing ℓ_1 reconstructs x exactly with overwhelming probability. In details, if the constant C is of the form $22(\delta+1)$ *in eq.* (2.2.4)*, then the probability of success exceeds* $1 - O(N^{-\delta})$.

- The first conclusion is that one suffers no information loss by measuring just about any set of K frequency coefficients.
- The second is that the signal x can be exactly recovered by minimizing a convex functional which does not assume any knowledge about the number of *nonzero coordinates of x*, *their locations*, and *their amplitudes* which we assume are all completely unknown a priori.

Theminimum number of samples needed for exact reconstruction by any method, no matter how intractable, must be about $S \log N$. ***Hence,*** the theorem theorem 2.2.1 is tight and ℓ_1-minimization succeeds nearly as soon as there is any hope to succeed by any algorithm.

By reversing the roles of time and frequency in the above example, we can recast theorem 2.2.1 as a new nonlinear sampling theorem. Thus, a nonlinear analog to Nyquist/Shannon: a signal can be reconstructed with *arbitrary and unknown* frequency support of size B from about $B \log N$ *arbitrarily chosen* samples in the time domain.

Finally, Fourier sampling theorem would be considered as a special instance of much more general statements. *As a matter of fact*[3], the results extend to a variety of other setups and higher dimensions. For instance, [8] shows how one can reconstruct a piecewise constant (on or two-dimensional) object from incomplete frequency samples provided that the number of jumps (discontinuities) obeys the condition above by minimizing other convex functionals such as the total variation.

2.2.2 Background

- Santosa and Symes [5] had suggested the minimization of ℓ_1-norms to recover sparse spike trains, see also [12, 33] for early results.
- In the last four years or so, a series of papers [2, 34–38] explained why ℓ_1 could recover sparse signals in some special setups.
- Important connections with the literature of theoretical computer science are pointed out. Inspired by [39], Gilbert and her colleagues have shown that one could recover an S-sparse signal with probability exceeding $1-\delta$ from $S \cdot poly(\log N, \log \delta)$ frequency samples placed on special *equispaced*[4] grids [11]. The algorithms they use are not based on optimization but rather on ideas from the theory of computer science such as *isolation* and *group testing*.
- Other points of connection include situations in which the set of spikes are spread out in a somewhat even manner in the time domain [10, 33].

[2] (P_1 can even be recast as a linear program [20, 32])
[3] 事实上
[4] 平均间隔；均布

2.2.3 Undersampling Structured Signals

The previous example showed that the structural content of the signal allows a drastic “undersampling”of the Fourier transform while still retaining enough information for exact recovery. ***Questions:*** to what extent can one recover a compressible signal from just a few measurements? What are good sensing mechanisms? Does *all this extend to object that are perhaps not sparse but well-approximated by sparse signals*? In the remainder of this paper, some answers to these fundamental questions are provided.

2.3 The Mathematics Of Compressive Sampling

2.3.1 Sparsity And Incoherence

As eq. (1.2.1) described, in practical instances, the vector x may be the coefficients of a signal $f \in \mathbb{R}^N$ in an orthonormal basis Ψ. For example, on might choose to expand the signal as a superposition of spikes (the canonical basis of $\mathbb{R}^N$), sinusoids, B-splines, wavelets, and so on. It is *not* important to restrict to orthogonal expansions as the theory and practice of compressive sampling accommodates other types of expansions. For example, x might be the coefficients of a digital image in a tight-frame of curvelets [40]. To keep on using convenient matrix notations, one can write the decomposition eq. (1.2.1) as

$$x = \Psi f$$

, where Ψ is the N by N matrix with the waveforms ψ_i as rows or equivalently,

$$f = \Phi^* x$$

.

A signal f is sparse in the Ψ-domain if the coefficient sequence is supported on a small set and compressible if the sequence is concentrated near a small set. Suppose we have available undersampled data about f of the same form as before

$$y = \Phi f.$$

Expressed in a different way, we collect partial information about x via $y = \Phi' x$ where $\Phi' = \Phi\Psi^*$. In this setup, one would recover f by finding - among all coefficient sequences consistent with the data - the decomposition with minimum ℓ_1-norm

$$\min \|\tilde{x}\|_{\ell_1} \qquad \text{such that} \qquad \Phi'\tilde{x} = y.$$

Of course, this is the same problem as eq. (2.2.3), which justifies the abstract and general treatment.

The key concept underlying the theory of compressive sampling is a kind of ***uncertainty relation***.

2.3.2 Recovery of Sparse Signals (sections 1.3.1 and 5.2.4)

About δ is in section 5.2.4.

The notion of *uniform uncertainty principle* (UUP)[5] was introduced in [31] and refined in [13].

UUP essentially states that $K \times N$ sensing matrix Φ obeys a “***restricted isometry hypothesis***”[6].

Let Φ_T $T \subset \{1, \cdots, N\}$ be the $K \times |T|$ submatrix obtained by extracting the columns of Φ corresponding to the indices in T; then [13] defines the S-restricted isometry constant δ_S of Φ which is the smallest quantity such that

$$(1-\delta_S)\|c\|_{\ell_2}^2 \le \|\Phi_T c\|_{\ell_2}^2 \le (1+\delta_S)\|c\|_{\ell_2}^2 \tag{2.3.1}$$

eq. (1.3.2) for all subsets T with $|T| \le S$ and coefficient sequences $(c_j)_{j\in T}$.

[5] 一致不确定原则
[6] 限制等距假设

This property essentially requires that *every set of columns with cardinality less than S approximately behaves like an orthonormal system*.

If the columns of the sensing matrix Φ are *approximately orthogonal*, then *the exact recovery phenomenon occurs.*

Theorem 2.3.1. *[13] Assume that x is S-sparse and suppose that $\delta_{2S} + \delta_{3S} < 1$ or, better, $\delta_{2S} + \theta_{S,2S} < 1$. Then the solution x^* to eq.* (2.2.3)[7] *is exact, $x^* = x$.*

In short, if the UUP holds at about the level S, the minimum ℓ_1-norm reconstruction is provably exact. The first thing one should notice when comparing this result with the Fourier sampling theorem is that it is deterministic in the sense that it *does not involve any probabilities*. It is also universal in that *all* sufficiently sparse vectors are exactly reconstructed from Φx.

In theorem 2.3.1, the version $\delta_{2S} + \theta_{S,2S} < 1$ is better, which is established in [17]. The number $\theta_{S,S'}$ for $S + S' \leq N$ is called the *S, S'-restricted orthogonality constants*[8] and is the smallest quantity such that

$$|\langle \Phi_T c, \Phi_{T'} c' \rangle| \leq \theta_{S,S'} \cdot \|c\|_{\ell_2} \|c'\|_{\ell_2} \tag{2.3.2}$$

holds for all *disjoint* sets $T, T' \subseteq \{1, \cdots, N\}$ of cardinality $|T| \leq S$ and $|T'| \leq S'$. Thus $\theta_{S\,S'}$ is the cosine of the smallest angle between the two subspaces spanned by the columns in T and T'. Small values of restricted orthogonality constants indicate that disjoint subsets of covariates span nearly orthogonal subspaces. The condition $\delta_{2S} + \theta_{S,2S} < 1$ is better than $\delta_{2S} + \delta_{3S} < 1$ since it is not hard to see that $\delta_{S+S'} - \delta_{S'} \leq \theta_{S,S'} \leq \delta_{S+S'}$ for $S' \geq S$ [Lemma 1.1 in [13]]

Suppose that $\delta_{2S} = 1$ which may indicate that there is a matrix $\Phi_{T_1 \cup T_2}$ with $2S$ columns ($|T_1| = S, |T_2| = S$) that is *rank-deficient*[9]. If this is the case, then there is a pair (x_1, x_2) of nonvanishing vectors with x_1 supported on T_1 and x_2 supported on T_2 obeying

$$\Phi(x_1 - x_2) = 0 \qquad \Leftrightarrow \qquad \Phi x_1 = \Phi x_2.$$

2.3.3 Recovery of Compressible Signals

[7] similar to eq. (1.3.3)
[8] 限制正交常量?
[9] 秩亏（降秩）

Chapter 3

“Compressed Sensing” Notes

3.1 abstract

假设 x 是 $\mathbf{R}^m$（一个数字图像或者信号）中的一个位置向量；我们计划测量 x 中的 n 个线性方程，然后重构它。如果 x 可被一个已知的变换编码进行压缩，并且可以通过本文定义的非线性过程进行重构，测量数量 n 可以远小于其实际大小 m。这样，相比常用的采样方法，确定的含有 m 像素的自然图像，仅需要 $n = O(m^{1/4} \log^{2/5}(m))$ 非自适应非像素采样就可以进行可靠恢复。

更明确的，假设 x 在一些正交基 (例如，小波，傅里叶) 和紧框架[1](curvelet, Garbor) 存在一个稀疏表示 —所以系数属于一个 ℓ_p 球，$0 < p \leq 1$。该扩展中，最重要的 N 个系数，在 ℓ_2

[1]tight frame

Chapter 4

Atomic Decomposition by Basis Pursuit

4.1 Abstract

The time-frequency and time-scale communities have recently developed a large number of overcomplete waveform dictionaries—stationary wavelets, wavelet packets, cosine packets, chirplets, and warplets, to name a few. Decomposition into overcomplete systems is not unique, and several methods for decomposition have been proposed, including the method of frames (MOF), matching pursuit (MP), and, for special dictionaries, the best orthogonalbasis (BOB). Basis pursuit (BP) is a principle for decomposing a signal into an "optimal"superposition of dictionary elements, where optimal means having the smallest l1 norm of coefficients among all such decompositions. We give examples exhibiting several advantages over MOF, MP, and BOB, including better sparsity and superresolution. BP has interesting relations to ideas in areas as diverse as ill-posed problems, abstract harmonic analysis, total variation denoising, and multiscale edge denoising. BP in highly overcomplete dictionaries leads to large-scale optimization problems. With signals of length 8192 and a wavelet packet dictionary, one gets an equivalent linear program of size 8192 by 212,992. Such problems can be attacked successfully only because of recent advances in linear and quadratic programming by interior-point methods. We obtain reasonable success with a primal-dual logarithmic barrier method and conjugategradient solver.

近年来，时 -频和时序共性已经发展出了很大数量的过完备波形字典 —稳定小波，小波包，余弦小波，线性调频小波，warplets, 等等还有许多。分解为超完备系统并不是唯一的方法，还有一些发表的分解方法，包括 *method of frames*(MOF), 匹配追踪 *Matching pursuit*(MP), 以及，对于特定字典的最佳正交基 *best orthogonal basis*(BOB).

将信号分解为一个"最优"字典成分叠加的原理是基追踪 (basis pursuit(BP))，其中"最优"指的是这些分解部分之间的系数存在最小 ℓ^1 泛数。本文举例列出了该方法相比 MOF,MP 和 BOB 等方法的一些优势，包括更好的稀疏度和超分辨率。BP 与一些领域的思想有着令人瞩目的关联，诸如不适定问题，抽象调和分析，总变差去噪和多尺度边缘去噪。

高度过完备字典中的 BP 引出了大规模优化问题。

Chapter 5

Decoding by Linear Programming

5.1 Abstract

本文研究针对真实数据输入/输出的自然错误纠正问题。我们希望从损坏的测量 $y = Af + e$ 中恢复输入向量 $f \in \mathbf{R}^n$。这里，A 是 $m \times n$（编码）矩阵，e 是一个任意且未知的误差向量。是否有可能从 y 中恢复确切的 f.

我们证明了在合适情况下的编码矩阵 A, 输入的 f 是 ℓ_1-最小化问题 ($\|x\|_{\ell_1} := \sum_i |x_i|$) 的唯一解

$$\min_{g \in \mathbf{R}^n} \|y - Ag\|_{\ell_1}$$

假设误差的支持向量[1]不是太大，$\|e\|_{\ell_0} := |\{i : e_i \neq 0\}| \leq \rho \cdot m$ for some $\rho > 0$。简言之，精确恢复 f 可以看作解决一个简单的凸优化问题（其可被重述为线性规划）。另外，数值实验表明了数据恢复过程的效果非常好；即使在数据的一个重要组成部分被破坏的情况下，f 仍可以被精确恢复。

本文工作相关于线性方程的极大欠定系统稀疏解的寻找问题。还与从高度不完整测量中恢复数据问题有着密切联系。实际上，本文研究结论是对我们之前工作的改进。最后，基于 ℓ_1 被称为一致不确定性[2]的一个重要性质，将在本文进行详述。

Index Terms—**Basic pursuit, decoding of (random) linear codes, duality in optimization, Gaussian random matrices, ℓ_1 minimization, linear codes, linear programming, principal angles, restricted orthonormality, singular values of random matrices, sparse solutions to underdetermined systems.**

5.2 Introduction

5.2.1 Decoding of Linear Codes

The linear code $a_1, \cdots, a_n \in \mathbf{R}^m$—the columns of the matrix A. There is a clear distinction between the real-valued setting and the finite alphabet one which is more common in the information theory literature.

Given $f \in \mathbf{R}^n$ (the **'plaintext'**) we can then generate a vector Af in $\mathbf{R}^m$(the "ciphertext"); if A has full rank, then one can clearly recover the plaintext f from the ciphertext Af. If A has full rank, then the plaintext f can be clearly recovered from the ciphertext Af. But with an arbitrary error vector $e \in \mathbf{R}^m$, can one recover f exactly from the given matrix A and $Af + e$?

The accurate decoding is impossible when the size of the support of the error vector is greater or equal to a half of that of the output Af. Therefore, a common assumption in the literature is to assume that only a small fraction of the entries are actually damaged

$$\|e\|_{\ell_0} := |i : e_i \neq 0| \leq \rho \cdot m. \tag{5.2.1}$$

For which values of ρ can we hope to reconstruct e with practical algorithms? That is, with algorithms whose complexity is at most polynomial in the length m of th code A?[3]

[1] 原文是 "the support of the vector"

[2] uniform uncertainty principle

[3] 为什么这里说：码字 A 的 m 长，复杂度最多的多项式的算法?

To reconstruct f, note that it is obviously sufficient to reconstruct the vector e since knowledge of $Af + e$ together with e gives Af, and consequently f, since A is full rank.

Approaches:

> Construct a matrix which annihilates the $m \times n$ matrix A on the left, such that $FA = 0$. This can be done in an obvious fashion by taking a matrix F whose kernel is the range of A in $\mathbf{R}^m$, which is an n-dimensional subspace (e.g., F could be the orthogonal projection onto the cokernel of A). We then apply F to the output $y = Af + e$ and obtain
>
> $$\tilde{y} = F(Af + e) = Fe \tag{5.2.2}$$
>
> since $FA = 0$.
>
> Therefore, the decoding problem is reduced to that of reconstructing a *sparse* vector e from the observations Fe (by sparse, we mean that only a fraction of the entries of e are nonzero). Therefore, *the overarching theme is that of the sparse reconstruction problem* [4].

5.2.2 Sparse Solutions to Underdetermined Systems

Finding sparse solutions to underdetermined systems of linear equations is in general NP-hard[5]. For example, the sparsest solution is given by

$$(P_0) \qquad \min d \in \mathbf{R}^m \|d\|_{\ell_0} \qquad \text{subject to} \qquad Fd = \tilde{y}(= Fe). \tag{5.2.3}$$

Solving this problem essentially requires exhaustive searches over all subjects of columns of F, a procedure which clearly is combinatorial in nature and has *exponential complexity.*

Formally, given an integer $p \times m$ matrix F and an integer vector b, the problem of deciding whether here is a vector with rational entries such that $Fd = b$, and with fewer than a fixed number of nonzero entries is NP-complete. In fact, the ℓ_0-minimization problem contains the subset-sum problem. Assume, for instance, that $m = 2k$ and $p = k + 1$, and consider the following set of $2k$ vectors in $\mathbf{R}^{k+1}$:

$$\begin{array}{cccc} e_1 & e_2 & \cdots & e_k \\ e_1 + a_1 e_{k+1} & e_2 + a_2 e_{k+1} & \cdots & e_k + a_k e_{k+1} \end{array}$$

where $e_1, \cdots, e_{k+1}$ are the usual basis vectors and $a_1, \cdots, a_k$ are integers. Now let α be another integer and consider the problem of deciding whether

$$e_1 + \cdots + e_k + \alpha e_{k+1}$$

is a k-sparse linear combination of the above $2k$ vectors ($k - 1$-sparse is impossible). This is exactly the subset-sum problem, i.e., whether one can write α as a sum of subset of $a_1, \cdots, a_k$. It is well known that this is an NP-complete problem.

A frequently discussed approach considers a similar (eq. (5.2.3)) program in the ℓ_1-norm which goes by the name of *basis pursuit* [20]

$$(P_1) \qquad \min x \in \mathbf{R}^m \|d\|_{\ell_1}, \qquad Fd = \tilde{y} \tag{5.2.4}$$

where we recall that $\|d\|_{\ell_q} = \sum_{i=1}^m |d_i|$. The ℓ_1-norm is convex. eq. (5.2.4) can be recast as a linear program (LP).

Motivated by the problem of finding sparse decompositions of special signals in the field of mathematical signal processing and following upon the *ground breaking*[6] work of [2], a series of beautiful articles [34, 35, 37, 41] showed exact equivalence between the two programs (eq. (5.2.3)) and (eq. (5.2.4)).

> In a nutshell, this work shows that for $m/2$ by m matrix F obtained by concatenation of two orthonormal bases, the solution to both eqs. (5.2.3) and (5.2.4) are unique and identical provided that in the most favorable case, the vector e has at most $0.914\sqrt{m/2}$ nonzero entries.

[4] 因此，最重要的主题是稀疏重构问题
[5] 非确定性多项式（non-deterministic polynomial，缩写 NP）
[6] 开创性的

This is of little practical use here since we are interested in procedures that might recover a signal when a constant fraction of the output is unreliable.

Using very different ideas and together with Romberg [8], the authors proved that the equivalence *holds* with overwhelming probability for various types of random matrices provided that the number of nonzero entries in the vector e be of the order of $m/\log m$ [31, 42]. In the special case where F is an $m/2$ by m random matrix with independent standard normal entries, [43] proved that the number of nonzero entries may be as large as $\rho \cdot m$, where $\rho > 0$ is some very small and unspecified positive constant independent of m.

5.2.3 Innovations

This paper introduces the concept of a restrictedly almost orthonormal system—a collection of vectors which behaves like an almost orthonormal system but only for sparse linear combinations. Thinking about these vectors as the columns of the matrix F, the authors showed that this condition allows for the exact reconstruction of sparse linear combination of these vectors. The results are significantly different than those mentioned above as they are deterministic and do not involve any kind of randomization, although *they can be specialized to random matrices.*

> For instance, A Gaussian matrix with independent entries sampled from the standard normal distribution is restrictedly almost orthonormal with overwhelming probability, and that minimizing the ℓ_1-norm recovers sparse decompositions with a number of **nonzero entries of size** $\rho_0 \cdot m$; the *numerical values for* ρ_0 should be given.

The connection with sparse solutions to underdetermined systems of linear equations is presented. The more direct approach:

> to recover f from corrupted data $y = Af + e$, solving the following ℓ_1-minimization problem is considered:

$$(P_1') \qquad \min_{g \in \mathbf{R}^n} \|y - Ag\|_{\ell_1}. \tag{5.2.5}$$

Now f is the unique solution of (P_1') if and only if e is the unique solution of (P_1). In other words, (P_1)eq. (5.2.4) and (P_1')eq. (5.2.5) are equivalent programs. Since $y = Af + e$, g can be decomposed as $g = f + h$ so that

$$(P_1') \quad \Leftrightarrow \quad \min_{h \in \mathbf{R}^m} \|x\|_{\ell_1}.$$

On the other hand, the constraint $Fx = Fe$ means that $x = e - Ah$ for some $h \in \mathbf{R}^m$

$$\begin{aligned} (P_1) \qquad &\Leftrightarrow \qquad \min_{h \in \mathbf{R}^m} \|x\|_{\ell_1}, \qquad x = e - Ah \\ &\Leftrightarrow \qquad \min_{h \in \mathbf{R}^m} \|e - Ah\|_{\ell_1} \end{aligned}$$

which proves the claim.

The program (P_1')eq. (5.2.5) may also be re-expressed as an LP. Indeed, the ℓ_1-minimization problem is equivalent to

$$\min 1^T t, \qquad -t \le y - Ag \le t \tag{5.2.6}$$

where the optimization variables are $t \in R^m$ and $g \in \mathbf{R}^n$ (as is standard, the generalized vector inequality $x \le y$ means that $x_i \le y_i$ for all i). As a result, (P_1') is and LP with inequality constraints and can be solved efficiently using standard optimization algorithms [44].

5.2.4 Restricted isometries (sections 1.3.1 and 2.3.2)

Here denote by $(v_j)_{j \in J} \in R^p$ the columns of the matrix F and by H the linear subspace spanned by these vectors. For any $T \subseteq J$, F_T is considered as ateh submatrix with column indices $j \in T$ so that

$$F_T c = \sum_{j \in T} c_j v_j \in H.$$

To introduce the notion of *almost orthonormal system*, we first observe that *if the columns of F are sufficiently "degenerate"*, the recovery problem cannot be solved. In particular, if there exists a nontrivial sparse linear combination $\sum_{j\in T} c_j v_j = 0$ of the v_j which sums to zero, and $T = T_1 \cup T_2$ is any prtition of T into two disjoint sets, then the vector y

$$y := \sum j \in T_1 c_j v_j = \sum j \in T_2 (-c_j) v_j$$

has two distinct sparse representations. On the other hand, linear dependencies $\sum_{j\in J} c_j v_j = 0$ which involve a large number of nonzero coefficients c_j, as *opposed to a sparse set of coefficients*, **do not** present an obvious obstruction to sparse recovery. *At the other extreme*, if the $(v_j)_{j\in J}$ are an orthonormal system, then the recovery problem is easily solved by setting

$$c_j = \langle f, v_j \rangle .$$

The main result of this paper is that if a "restricted orthonormality hypothesis" is imposed, which is far weaker than assuming orthonormality, then (P_1)eq. (5.2.4) solves the recovery problem, even if the $(v_j)_{j\in J}$ are highly linearly dependent (for instance, it is possible for $m := |J|$ to be much larger than the dimension of the span of the v_j's). The following definition introduced make this quantitative

Definition 5.2.1 (**Restricted Isometry Constants**)**.** *Let F be the matrix with the finite collection of vectors $(v_j)_{j\in J} \in \mathbf{R}^p$ as columns. For every integer $1 \le S \le |J|$, we define the* ***S-restricted isometry constants δ_S to be the smallest quantity*** *such that F_T obeys*

$$(1-\delta_S)\|c\|^2 \le \|F_T c\|^2 \le (1+\delta_S)\|c\|^2 \tag{5.2.7}$$

for all subsets $T \subset J$ of cardinality at most S, and all real coefficients $(c_j)_{j\in T}$. Similarly, we define the S,S'-restricted orthogonality constraints $\theta_{S,S'}$ for $S + S' \le |J|$ to be the smallest quantity such that

$$|\langle F_T c, F_{T'} c' \rangle| \le \theta_{S,S'} \cdot \|c\| \|c'\| \tag{5.2.8}$$

holds for all disjoint sets $T, T' \subseteq J$ of cardinality $|T| \le S$ and $|T'| \le S'$.

> **δ is also described in section 2.3.2.**
>
> The number δ_S and $\theta_{S,S'}$ measure **how close the vectors v_j are to behaving like an orthonormal system**, but only when restricting attention to sparse linear combinations involving no more than S vectors.
>
> δ_S and $\theta_{S,S'}$ 测量了向量 v_j 表现为一个标准正交系统的近似程度，但仅关注在包含不超过 S 个向量的稀疏线性组合情况中。
>
> These numbers are clearly nondecreasing in S, S'. For $S = 1$, the value δ_1 only conveys magnitude information about the vectors v_j; 这两个数值在 S, S' 中显然是非减的。对于 $S = 1$，δ_1 仅表达向量 v_j 的大小；indeed δ_1 is the best constant such that
>
> $$1-\delta_1 \le \|v_j\|^2 \le 1+\delta_1, \qquad \text{for all } j \in J. \tag{5.2.9}$$

$\delta_1 = 0$ if and only if all of the $v_j's$ have unit length. Higher δ_S control the orthogonality numbers $\theta_{S,S'}$.

Lemma 5.2.1. *We have $\theta_{S,S'} \le \delta_{S+S'} \le \theta_{S,S'} + \max(\delta_S, \delta_{S'})$ for all S, S'.*

To see the relevance of the δ_S to the sparse recovery problem, consider the following simple observation.

Lemma 5.2.2. *Suppose that $S \ge 1$ is such that $\delta_{2S} < 1$, and let $T \subset T$ be such that $|T| \le S$. Let $f := Fc$ where c is an arbitrary vector supported on T. Then c*

$$(P_0) \qquad \min \|d\|_{\ell_0}, \qquad Fd = f$$

so that c can be reconstructed from knowledge of the vector f (and the $v_j's$).

Proof. There is a unique c with $\|c\|_{\ell_0} \leq S$ and obeying $f = \sum_j c_j v_j$. Suppose f had two distinct sparse representations $f = Fc = Fc'$ where c and c' were supported on sets obeying $|T|, |T'| \leq S$. Then

$$F(c - c') = 0.$$

□

By construction $c - c'$ is supported on $T \cup T'$ of size less or equal to $2S$. Taking norms of both sides and applying eq. (5.2.7) and the hypothesis $\delta_{2S} < 1$ we conclude that $\|c - c'\|^2 = 0$, contradicting the hypothesis that the two representations were distinct.

5.2.5 Main Results

Previous lemma is an abstract existence argument which shows what might theoretically be possible, but does not supply any efficient algorithm to recover T and c_j from f and $(v_j)_{j \in J}$ other than by brute-force search—as discussed earlier. The main theorem shows that, by imposing slightly stronger conditions on δ_{2S}, the ℓ_1-minimization program (P_1) recovers f exactly.

Theorem 5.2.1. *Suppose that $S \geq 1$ is such that*

$$\delta_S + \theta_{S,S'} + \theta_{S,2S} < 1 \tag{5.2.10}$$

and let c be a real vector supported on a set $T \subset J$ obeying $|T| \leq S$. Put $f := Fc$. Then c is the unique minimizer to

$$(P_1) \qquad \min \|d\|_{\ell_1}, \qquad Fd = f.$$

lemma 5.2.1 implies $\delta_{2S} < 1$, and is in turn implied by $\delta_S + \delta_{2S} + \delta_{3S} < 1$. Thus, the condition eq. (5.2.10) is roughly "three times as strict" as the condition required for lemma 5.2.2.

theorem 5.2.1 is inspired by [31], see also [8, 42], but the results in this paper are *deterministic*, and thus do not have ***a nonzero probability of failure***, provided of course on can ensure that the system $(v_j)_{j \in J}$ verifies the condition eq. (5.2.10). By virtue of the previous discussion, we have the companion result.

Theorem 5.2.2. *Suppose F is such that $FA = 0$ and let $S \geq 1$ be a number obeying the hypothesis of theorem 5.2.1. Set $y = Af + e$, where e is a real vector supported on a set of size at most S. Then f is the unique minimizer to*

$$(P_1') \qquad \min_{g \in \mathbf{R}^n} \|y - Ag\|_{\ell_1}.$$

$\boxed{\text{test hello } \alpha}$

Chapter 6

Stable Signal Recovery from Incomplete and Incaccurate measurements

Bibliography

[1] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *Ieee Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[2] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, 2001.

[3] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Applied and Computational Harmonic Analysis*, vol. 10, no. 1, pp. 27–44, 2001.

[4] J. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophys. Mag.*, vol. 38, no. 5, pp. 826–844, Oct 1973.

[5] F. Santosa and W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM J. Sci. Statist. Comput.*, vol. 7, no. 4, pp. 1307–1330, 1986.

[6] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.

[7] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[8] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[9] F. Ping and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, vol. 3, May 1996, Conference Proceedings, pp. 1688–1691 vol. 3.

[10] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *Ieee Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[11] A. C. Gilbert, S. Muthukrishnan, and M. Strauss, "Improved time bounds for near-optimal sparse fourier representations," vol. 5914, 2005, Conference Proceedings, pp. 59 141A–59 141A–15.

[12] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Mathematical Analysis*, vol. 49, no. 3, pp. 906–931, 1989.

[13] E. J. Candes and T. Tao, "Decoding by linear programming," *Ieee Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[14] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug 2006.

[15] E. Candès, "Lectures on compressive sampling and frontiers in signal processing," The Institute for Mathematics and its Applications. University of Minnesota, Tech. Rep., June 2007.

[16] A. Cohen, W. Dahmen, and R. Devore, “Compressed sensing and best k-term approximation,” *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.

[17] E. Candes and T. Tao, “The dantzig selector: Statistical estimation when p is much larger than n,” *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, Dec 2007.

[18] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” *Ieee Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, Sep 2006.

[19] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1994.

[20] S. S. B. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *Siam Review*, vol. 43, no. 1, pp. 129–159, Mar 2001.

[21] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008/12/01 2008.

[22] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, “Uniform uncertainty principle for bernoulli and subgaussian ensembles,” *Constructive Approximation*, vol. 28, no. 3, pp. 277–289, 2008/12/01 2008.

[23] D. L. Donoho, “Compressed sensing,” *Ieee Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr 2006.

[24] M. Rudelson and R. Vershynin, “On sparse reconstruction from fourier and gaussian measurements,” *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, Aug 2008.

[25] R. Blahut, *Algebraic Codes for Data Transmission.* Cambridge, U.K.: Cambridge Univ. Press, 2003.

[26] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, “Distributed compressed sensing,” *preprint*, 2005.

[27] J. M. P. M. Lustig, D. L. Donoho, “Rapid mr imaging with compressed sensing and randomly under-sampled 3dft trajectories,” *in Proc. 14th*, 2006.

[28] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, K. Kelly, and R. Baraniuk, “A compressed sensing camera: New theory and an implementation using digital micromirrors,” *Proc. Computational Imaging IV at SPIE Electronic Imaging, San Jose*, 2006.

[29] E. J. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted l(1) minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, Dec 2008.

[30] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, “Data compression and harmonic analysis,” *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2435–2476, 1998.

[31] E. J. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *Ieee Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec 2006.

[32] P. Bloomfield and W. L. Steiger, “Least absolute deviations. theory, applications, and algorithms,” *Progress in probability and statistics*, 1983.

[33] D. L. Donoho and B. F. Logan, “Signal recovery and the large sieve,” *SIAM Journal on Applied Mathematics*, vol. 52, no. 2, pp. 577–591, 1992.

[34] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l(1) minimization,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, Mar 2003.

[35] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *Ieee Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, Sep 2002.

[36] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *Information Theory, IEEE Transactions on*, vol. 49, no. 6, pp. 1579–1581, 2003.

[37] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *Ieee Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec 2003.

[38] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *Ieee Transactions on Information Theory*, vol. 50, no. 6, pp. 1341–1344, Jun 2004.

[39] Y. Mansour, "Randomized interpolation and approximation of sparse polynomials," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 357–368, 1995.

[40] E. J. Candes and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise c-2 singularities," *Communications on Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, Feb 2004.

[41] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.

[42] E. J. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics*, vol. 6, no. 2, pp. 227–254, 2006/04/01 2006.

[43] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l(1)-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, Jun 2006.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.