压缩感知*

许志强†
中国科学院数学与系统科学研究院,
计算数学与科学工程计算研究所,
科学与工程计算国家重点实验室, 100190, 北京

2012 年 1 月 12 日

摘要

压缩感知是近来国际上热门的研究方向. 其在信号处理中具有很好的应用前景. 此外, 它与逼近论、最优化、随机矩阵及离散几何等领域密切相关, 由此产生了一些漂亮的数学结果. 本文综述压缩感知一些基本结果并介绍最新进展. 主要包括 RIP 矩阵编码与 ℓ_1 解码的性能, RIP 矩阵的构造, Gelfand 宽度, 个例最优性及 OMP 解码等.

1 引言

现实世界中, 人们经常需要对信号进行观测, 例如医学图像成像、CT 断层扫描等, 以期通过观测信息对原始的信号进行重建. 由于计算机的离散化存储, 我们可将需重建的信号 x 抽象为一 N 维向量, 可将对信号 x 的观测抽象为用一 $n \times N$ 的矩阵 Φ 与信号 x 进行乘积. 例如在 CT 扫描中, 矩阵 Φ 通常选择为离散 Fourier 矩阵. 那么, 我们所观测的信息为

$$y = \Phi x. \tag{1}$$

人们自然而问: **为重建信号 x, 至少需要多少次观测?** 由线性代数知识可知, 为使方程组 (1) 的解存在且唯一, 我们须选择 $n \geq N$. 也就是说, 我们需要至少进行 $n = N$ 次观测. 然而, 现实世界中的自然信号通常具有一定规律性. 对这种规律性, 一种常用的刻画方式是自然信号在一组基底表示下是稀疏的. 这里的“稀疏”是指它们用一组基底展开后, 大多数系数为 0, 或者绝对值较小. 例如, 自然图像用小波基底展开后, 一般而言, 其展开系数大多

*国家自然科学基金 (11171336) 及创新群体 (11021101) 资助.
†Email: xuzq@lsec.cc.ac.cn

数绝对值较小. 这也就是图像能够进行压缩的原理. 然而, 这同时为人们减少观测次数 n 从理论上提供了可能性. 因而, 压缩感知的主要任务为: **对尽量小的 n, 设计 $n\times N$ 观测矩阵 Φ, 以及通过 Φx 快速恢复 x 的算法**. 所以, 压缩感知的研究主要分为两方面：矩阵 Φ 的设计; 与反求信号 x 的算法.

本文主要介绍压缩感知的一些基本结果. 在每节里, 我们采用注记的方式介绍当前的一些研究进展及研究问题, 同时提供与之相关的参考文献, 以使感兴趣的读者可进一步探索. 本文组织结构如下: 第 2 节中我们介绍了稀疏信号精确恢复的编码、解码方法. 特别是, 我们将介绍矩阵的零空间性质, 及 RIP 矩阵编码与 ℓ_1 解码的性能. 我们在第 3 节中介绍 RIP 矩阵的构造方法, 包括随机矩阵、结构随机矩阵及确定性矩阵. 在第 4 节中, 为理解最优编码、解码对的性能, 我们介绍了 Gelfand 宽度与编码、解码对性能的关联. 我们在第 5 节中介绍了编码、解码对在不同范数意义下的个例最优性. 最后一节简要介绍实现解码的算法.

2 稀疏信号的恢复

为方便介绍压缩感知理论, 我们将信号的稀疏性简单理解为信号中非 0 元素数目较少. 我们所指的信号即为一向量 $x\in\mathbb{R}^N$. 我们用 Σ_s 表示 s-稀疏向量集合, 即

$$\Sigma_s:=\{x\in\mathbb{R}^N:\|x\|_0\le s\},$$

这里 $\|x\|_0$ 表示 x 中的非 0 元素数目. 所谓对信号 $x_0\in\mathbb{R}^N$ 编码, 即指用一 $n\times N$ 的矩阵 Φ 与 $x_0\in\mathbb{R}^N$ 进行乘积, 那么我们得到

$$y=\Phi x_0.$$

此处, $y\in\mathbb{R}^n$ 即为我们所观测到的关于 x_0 的信息. 所谓解码, 就是试图通过 y 反求 x_0, 也就是寻找一从 $\mathbb{R}^n$ 到 $\mathbb{R}^N$ 的映射, 我们将该映射记为 Δ. 我们用 $\Delta(y)$ 表示反求结果. 一般而言, 若 $n<N$, 则有无数个 $x\in\mathbb{R}^N$ 满足 $y=\Phi x$. 因而, 只有借助信号稀疏性的特征, 我们才有可能反求原始的信号 x_0.

那么, 给定一编码、解码对 (Φ,Δ), 我们关心其性能, 即:

$$\|x_0-\Delta(\Phi x_0)\|_X,$$

此处 X 为一给定范数. 本文中, 我们通常选择 X 为 ℓ_p 范数, 并用下标 p 表示 ℓ_p 范数. 当 x_0 中非 0 元素数目较小的时候, 一种较为自然的解码 $\Delta_0(y)$ 是如下规划问题的解:

$$\begin{aligned} P_0:\quad \min_{x\in\mathbb{R}^N}\quad & \|x\|_0\\ \text{s.t.}\quad & \Phi x=y. \end{aligned}$$

这里, $\|x\|_0$ 表示 x 中非 0 元素的数目. 我们用符号 $\Delta_0(y)$ 表示 P_0 的解. 也就是说, $\Delta_0(y)$ 为在所有满足线性方程组 $\Phi x = y$ 的向量中, 选择非 0 元素数目最少的. 如果我们对矩阵 Φ 加些许限制, 由 P_0 定义的解码, 可精确恢复 s-稀疏向量:

定理 2.1 假定 $\Phi \in \mathbb{R}^{n\times N}$ 是一个任 $2s$ 列均线性无关的矩阵. 我们选择解码为 Δ_0, 那么, 对任意的 $x_0 \in \Sigma_s$,

$$\Delta_0(\Phi x_0) = x_0.$$

根据定理 2.1, 如果我们选取观测次数 $n = 2s$, 那么就存在一编码、解码对 (Φ, Δ) 使得对任意的 $x_0 \in \Sigma_s$, 均有 $\Delta(\Phi x_0) = x_0$. 这意味着: 如果我们希望恢复一个嵌入在 N 维空间的 s-稀疏向量, 那么 $2s$ 次观测次数就足够了.

但是, 问题 P_0 的求解是十分不平凡的. 事实上, P_0 是一个 NP 完全问题 [14, 31]. 那么, 我们能否找到一个更为有效的解码算法? 一个令人惊讶的事实是, 如果矩阵 Φ 满足一定条件, 那么回答则是肯定的, 但我们要在观测次数上付出些许代价. 我们现在将解码 $\Delta_1(y)$ 定义为如下问题的解:

$$P_1: \quad \begin{array}{ll} \min\limits_{x\in\mathbb{R}^N} & \|x\|_1 \\ \text{s.t.} & \Phi x = y. \end{array}$$

如所知, P_1 可转化为如下的线性规划问题：

$$P_2: \quad \begin{array}{ll} \min\limits_{t\in\mathbb{R}^N} & t_1 + t_2 + \cdots + t_N \\ \text{s.t.} & \Phi x = y \\ & -t_j \le x_j \le t_j,\ j = 1, \ldots, N \\ & t_j \ge 0,\ j = 1, \ldots, N. \end{array}$$

从一个简单的论证可看出, P_1 的解与 P_2 的解相同. 因此, 可找到有效的算法对 P_1 求解. 但是, 一个自然的问题是: P_1 的解与 P_0 的解等价吗? 或者是,

对什么样的观测矩阵 Φ, P_1 的解与 P_0 的解总一致?

为回答这一问题, 我们首先介绍矩阵的**零空间**性质. 零空间性质思想的主要出发点是：解码通常是从集合

$$\{x \in \mathbb{R}^N : \Phi x = y\}.$$

中按一定规则挑选我们需要的元素. 由线性代数知识可知, 解集 $\{x \in \mathbb{R}^N : \Phi x = y\}$ 可由原始的信号 x_0, 与矩阵的 Φ 的零空间

$$\text{Ker }\Phi \ := \ \{x \in \mathbb{R}^N : \Phi x = 0\}$$

所确定. 因此, 人们考虑通过刻画矩阵 Φ 的零空间, 从而给出 P_1 解与 P_0 解一致的充要条件. 我们首先介绍零空间性质的定义. 为描述方便, 我们介绍如下符号: 对于指标集 $T \subset \{1,\ldots,N\}$ 及向量 $v \in \mathbb{R}^N$, 我们将 v 中指标在 T 中的元素取出, 形成一个新的向量, 标记为 $v_T \in \mathbb{R}^{\#T}$. 我们用 T^c 表示 T 的补集. 类似的, 我们可定义矩阵 Φ_T.

定义 2.1 *我们称矩阵 Φ 满足 s-阶零空间性质, 如果对任意的 $v \in \mathrm{Ker}\,\Phi$, 均有*

$$\|v_T\|_1 < \|v_{T^c}\|_1, \quad \text{对任意的 } T \subset \{1,\ldots,N\},\ \#T = s.$$

直观上, 我们将零空间性质理解为 $\mathrm{Ker}\,\Phi$ 的非 0 元素较为均匀的分布, 不会明显的集中于某 s 个元素上. 采用零空间性质, 我们有

定理 2.2 *我们选择解码为 Δ_1. 那么, 对任意的 $x \in \Sigma_s$,*

$$\Delta_1(\Phi x) = x$$

如果和仅仅如果 Φ 满足 s-阶零空间性质.

注 2.1 *用类似于零空间性质的方式, 描述 Δ_1 解码可恢复 s-稀疏信号, 在人们研究 L_1 最佳逼近时就已经出现 (参考 [33]). 与定理 2.2 一致的形式首先出现在 [23]. 此外, 文 [18, 19] 也隐含了类似的结果.*

虽然可以用零空间性质给出 P_1 的解与 P_0 的解一致的充要条件. 但是, 零空间性质并不容易操作, 无论在理论还是计算方面. 也就是说, 给一个矩阵 Φ, 难以从理论上证明其是否满足零空间性质, 也不容易在计算机上快速验证. 因而, 人们考虑了另外一种刻画方式, 即是所谓的**矩阵 RIP 性质** (Restricted Isometry Property).

我们首先介绍 RIP 性质的定义 [9]. 我们说矩阵 Φ 满足 s-阶 RIP 性质, 如果存在常数 $\delta_s \in [0,1)$ 使得

$$(1-\delta_s)\|x\|^2 \le \|\Phi x\|^2 \le (1+\delta_s)\|x\|^2 \tag{2}$$

对任意的 $x \in \Sigma_s$ 成立. 实际上, (2) 等价于 Grammian 矩阵 $\Phi_T^\top \Phi_T$ 所有特征值位于区间 $[1-\delta_s, 1+\delta_s]$, 这里 $\#T \le s$. 我们称 δ_s 为 RIP 常数.

我们首先看一下如何从直观上理解 RIP 矩阵. 倘若 $\delta_s = 0$, 那么矩阵 Φ 为一标准正交矩阵. 因而也是一方阵. 然而在压缩感知中, 我们希望矩阵 Φ 是“扁” 的, 也就是 $n < N$, 同时保留类似于正交矩阵的特征. 因而, RIP 矩阵的定义 (2), 事实上刻画了矩阵 Φ 中任取 s 列所形成的 $n \times s$ 矩阵接近于正交矩阵的程度. RIP 常数 δ_s 越接近于 0, 其任取 s 列所形成的子矩阵也就越接近于正交. 从某种意义上来说, 性质也就越好.

下面定理给出了解码 Δ_1 能够精确恢复 s-稀疏信号的一个充分条件.

定理 2.3 ([6, 7]) *假定编码矩阵 Φ 满足 $2s$ 阶 RIP 性质，且 RIP 常数 $\delta_{2s}\le\sqrt{2}-1$. 我们选择解码 Δ_1. 那么，对任意的 $x\in\Sigma_s$，均有*

$$\Delta_1(\Phi x)=x.$$

上述定理表明, 我们可用 RIP 矩阵编码、Δ_1 解码精确恢复 s-稀疏信号. 然而, 现实世界中的信号并非严格稀疏的, 通常仅仅是近似稀疏. 对于此类信号, 如果我们仍然用 RIP 矩阵 Φ 进行编码, 选则解码为 Δ_1, 那么我们能较好的恢复非稀疏信号吗? 令人惊讶的是, 我们仍然能较好的完成任务. 为介绍这方面的结果, 我们首先介绍**最佳 s-项逼近误差**的概念. 给定范数 $\|\cdot\|_X$, 那么信号 $x\in\mathbb{R}^N$ 的在范数 $\|\cdot\|_X$ 意义下**最佳 s-项逼近误差**为

$$\sigma_s(x)_X:=\min_{z\in\Sigma_s}\|x-z\|_X.$$

对于 $K\subset\mathbb{R}^N$, 我们令

$$\sigma_s(K)_X:=\max_{x\in K}\sigma_s(x)_X.$$

下面定理指出, 对于一般信号, 我们采用 RIP 矩阵编码与用 Δ_1 作解码, 那么恢复效果可用 ℓ_1 范数意义下的最佳逼近误差刻画.

定理 2.4 *假定编码矩阵 Φ 满足 $2s$ 阶 RIP 性质，且 RIP 常数 $\delta_{2s}\le\sqrt{2}-1$. 我们选择解码为 Δ_1. 那么对任意的 $x\in\mathbb{R}^N$，我们有*

$$\|\Delta_1(\Phi x)-x\|_2\ \le\ C_0\frac{\sigma_s(x)_1}{\sqrt{s}},$$

此处 C_0 为一常数.

注 2.2 *定理 2.3 与定理 2.4 首先在[7] 中被证明. 但给出的 RIP 常数较为粗糙. 在 [6] 中，E. Candès 将RIP 常数改进为 $\delta_{2s}\le\sqrt{2}-1$. 仍有一些论文考虑改进定理 2.3 中的 RIP 常数 $\sqrt{2}-1$ [4, 21]. 特别是，Mo 和 Li 将该常数改进为 $\delta_{2s}<0.4931$ [30]. 此外，Davies 和 Gribonval 建构一个例子表明，如果 $\delta_{2s}\ge\frac{1}{\sqrt{2}}$，那么 Δ_1 解码不能恢复一些 s-稀疏信号. 注意到这些研究均是针对 δ_{2s}，也就是要求矩阵满足 $2s$ 阶 RIP 条件. 在 [5] 中，Cai, Wang 和 Xu 考虑了矩阵满足 s-阶 RIP 条件的情况，给出了 P_1 能恢复 s-稀疏信号的充分条件为 $\delta_s<0.307$. 此外，我们特别指出，借助离散几何中的多面体理论，在文 [16] 中，Donoho 给出了 Δ_1 解码能精确恢复 s-稀疏信号的充要条件.*

注 2.3 *对于 $0<p<1$，人们也考虑了如下定义的 Δ_p 解码:*

$$\min_{x\in\mathbb{R}^N}\quad\|x\|_p\qquad\text{s.t.}\quad\Phi x=y.$$

这里 $\|x\|_p := (\sum_{j=1}^N |x_j|^p)^{1/p}$. 事实上, 当 $0<p<1$, $\|\cdot\|_p$ 为一拟范数. 相比于 Δ_1 解码, Δ_p 解码所需观测次数较少, 但解码复杂度会有所增加*[39, 13]*.

注 2.4 *在本文中, 我们假定信号的稀疏性指非 0 元素较少. 但是, 很多应用问题里面, 信号是在一“字典”或者紧框架表示下是稀疏的. 对于此类情况, 文 [11] 进行了研究. 并将定理 2.4 进行了推广. 但是, 这个方向仍值得进一步深入探索.*

我们现在回到本文开始所提出的问题：

如果选择解码为 Δ_1, 为精确恢复所有 s-稀疏信号, 观察次数 n 最少应为多少?

文 [22] 给出了如下定理:

定理 2.5 *假定 $\Phi \in \mathbb{R}^{n\times N}$ 及解码为 Δ_1. 那么, 如果*

$$\Delta_1(\Phi x) = x, \qquad \text{对任意} \qquad x \in \Sigma_{2s},$$

则

$$n \geq c_1 s \log\left(\frac{N}{c_2 s}\right),$$

这里 $c_1 = \frac{1}{\log 9} \approx 0.455$ 且 $c_2 = 4$.

根据上述定理, 如果解码选择为 Δ_1, 那么我们至少需要进行 $n = \mathcal{O}(s\log\left(\frac{N}{s}\right))$ 次观测, 才能精确恢复 s-稀疏信号. 这个下界能够达到吗? 也就是说, 对于 $n = \mathcal{O}(s\log\left(\frac{N}{s}\right))$, 我们能否构造观测矩阵 $\Phi \in \mathbb{R}^{n\times N}$, 使得对任意 $x \in \Sigma_s$, 均有 $\Delta_1(\Phi x) = x$? 根据定理 2.3, 我们可将该问题归结为能否构造满足 s-阶 RIP 条件的矩阵 $\Phi \in \mathbb{R}^{n\times N}$ 使得 $n = \mathcal{O}(s\log\left(\frac{N}{s}\right))$? 我们将在下节回答该问题.

3 RIP 矩阵

根据定理 2.3 和定理 2.4, 为保证 ℓ_1 解码能恢复稀疏或者近似稀疏信号, 我们需要构造 RIP 矩阵. 我们希望对于给定的 $n, N \in \mathbb{Z}$, 构造一 $n\times N$ 的矩阵 Φ, 以使其对尽量大的 s 满足 s-阶 RIP 条件. 那么, 如何构造此类矩阵? 当前的主要构造方法有：随机矩阵、结构随机矩阵与确定性矩阵.

3.1 随机矩阵

我们考虑两类随机矩阵：Gaussian 随机矩阵与 Bernoulli 随机矩阵. 所谓 Gaussian 随机矩阵, 即指矩阵中的元素 $\phi_{i,j}$ 是独立的随机变量且服从如下分布:

$$\phi_{i,j} \sim \mathcal{N}(0, \frac{1}{n})$$

即服从期望为 0, 方差为 $\frac{1}{n}$ 的 Gaussian 分布. 所谓 Bernoulli 矩阵, 即指矩阵 Φ 中的元素以相同的概率取 $\frac{1}{\sqrt{n}}$ 或 $-\frac{1}{\sqrt{n}}$.

定理 3.1 *假定 $n \times N$的矩阵 Φ 是一个 Gaussian 或者 Bernoulli 随机矩阵. 那么, 当*

$$s \leq C_1 n/\log(N/s)$$

矩阵 Φ 是一个 s-阶 RIP 矩阵的概率不小于

$$1 - \exp(-C_2 n),$$

此处常数 C_1, C_2 仅仅依赖于 RIP 常数 δ.

注 3.1 *一个类似于定理 3.1 的结果最早由 Kashin 得到 [24]. 文 [10, 41] 中也给出了定理 3.1 的证明. 一个比较简单的证明方法是 [1] 中所介绍的. 此外, 文 [1] 也给出了 RIP 性质与 Johnson-Lindenstrauss 引理的关联.*

注 3.2 *定理 3.1 表明 Gaussian 随机矩阵或 Bernoulli 随机矩阵满足 $s = \mathcal{O}(n/\log(N/s))$ 阶的 RIP 性质. 根据逼近论中的宽度理论, 对于给定的 $n, N \in \mathbb{N}$, 这里的 s 已经达到了最佳阶.*

3.2 确定性矩阵

虽然随机矩阵能产生尺寸接近最优的 RIP 矩阵. 在工程实际中, 人们更希望构造一个确定性 RIP 矩阵. 因为确定性矩阵更利于工程设计, 此外, 从构造解码算法角度来看, 确定性矩阵利于降低内存、设计快速的恢复算法等. 然而, 现在仍然缺少令人满意的确定性 RIP 矩阵构造方法. 当前的构造方法主要是基于矩阵的**列相干性**.

假定矩阵 $\Phi = (a_1, \ldots, a_D) \in \mathbb{C}^{n\times N}$, 这里 $n \leq N$. 我们假定矩阵 Φ 中的列元素标准化, 即 $\|a_i\|_2 = 1$. 矩阵 Φ 的列相干性定义为

$$\mathcal{M}(\Phi) := \max_{i\neq j} |\langle a_i, a_j \rangle|.$$

下式给出了 $\mathcal{M}(\Phi)$ 的一个下界, 也称为Welch 界 [43]

$$\mathcal{M}(\Phi) \geq \sqrt{\frac{N-n}{(n-1)N}}. \tag{3}$$

当等号成立的时候, 我们称矩阵 Φ为**最优 Grassmannian 框架**. 文 [20] 中指出, 只有当 $N \leq n^2$, 等号才有可能成立 (参考 [40]).

下面定理显示了矩阵的列相干性与 RIP 性质之间的关联 [15, 2].

定理 3.2 *假定 $a_1, \ldots, a_N$ 是矩阵 Φ 的列元素且其列相干性为 μ. 那么, 矩阵 Φ 满足 RIP 常数为 $\delta_s = (s-1)\mu$ 的 s-阶 RIP 性质.*

人们能够构造出满足条件

$$\mu = \mathcal{O}\left(\frac{\log N}{\sqrt{n}\log n}\right)$$

的矩阵 (参考 [25, 15, 45]). 我们在此介绍作者在 [45] 中提出的一种构造方法, 其主要利用了数论中的 Weil 指数和定理 [42]:

定理 3.3 ([42]) *假定 p 是一个素数. 假定 $f(x) = m_1 x + \cdots + m_d x^d$, 且存在一个 j, $1 \leq j \leq d$, 使得 $p \nmid m_j$. 那么*

$$\left|\sum_{x=1}^{p} e^{\frac{2\pi i f(x)}{p}}\right| \leq (d-1)\sqrt{p}.$$

给定正整数 q 和 d, 我们下面构造一 $n \times N$ 矩阵 Φ, 这里 $n \geq 2q+1$ 为素数, 且 $N = (2q+1)^d$. 那么, 矩阵 Φ 的第 j 行定义为

$$\Phi_{j,\cdot} = \left[\frac{\exp(2\pi i \langle x_j, k\rangle)}{\sqrt{n}}\right]_{k\in[-q,q]^d} \qquad \in \qquad \mathbb{C}^{(2q+1)^d}, \tag{4}$$

这里

$$x_j = [j, j^2, \ldots, j^d]/n \mod 1.$$

下面的定理刻画了由 (4) 所定义的矩阵 Φ 的列相干性 [45].

定理 3.4 *给定正整数 q 和 d, 令 $n \geq 2q+1$ 为素数, 且 $N = (2q+1)^d$. 假设 $n \times N$ 矩阵 Φ 由 (4) 定义. 那么,*

$$\mathcal{M}(\Phi) \leq \frac{d-1}{\sqrt{n}}.$$

根据 Bertrand-Chebyshev 定理, 在区间 $[2q+1, 4q+2]$ 中必存在一素数. 因此, 我们可以假定 $n \leq 4q+2$. 那么, 对定理 3.4 中的 Φ, 我们有

$$\mathcal{M}(\Phi) \leq \frac{d-1}{\sqrt{n}} \leq 2\frac{\log N}{\sqrt{n}\log n}.$$

组合定理 3.2 和定理 3.4, 我们有

定理 3.5 *定理 3.4 中的 Φ 满足 $s = \mathcal{O}\left(\frac{\sqrt{n}}{d}\right)$ 阶 RIP 条件.*

在文 [45] 中, 作者也通过数值试验显示该确定性矩阵 Φ 与随机矩阵的编码效果基本一致. 但是, 根据定理 3.1, 随机矩阵能够满足 $s = \mathcal{O}(n/\log(N/s))$ 阶 RIP 性质. 这要优于由 (4) 所定义的确定性矩阵 Φ 的 $\mathcal{O}(\sqrt{n}/d)$.

当 $N \geq 2n$, 根据 Welch 界, 对任意的 $n \times N$ 矩阵 Φ,

$$\mu = \mathcal{M}(\Phi) \geq \frac{1}{\sqrt{2(n-1)}}.$$

因而

$$\frac{s-1}{\sqrt{2(n-1)}} \leq (s-1)\mu < 1.$$

我们由此得到, 矩阵 Φ 满足 $s \leq \sqrt{2n}$ 阶 RIP 条件. 这个界说明, 如果我们仅仅分析矩阵的列相干性, 难以论证确定性矩阵满足 $s = \mathcal{O}(n/\log(N/s))$ 阶 RIP 性质. 最近, 借助加性组合与解析数论的工具, Bourgain 等人证明了, 当 $d = 2$, 由 (4) 所定义的矩阵 Φ 满足 $s = n^{1/2+\epsilon_0}$ 阶 RIP 性质, 这里 ϵ_0 是一个充分小的正数. 这突破了由分析矩阵的列相干性所带来的 1/2 瓶颈 $s = \mathcal{O}(n^{1/2})$. 然而, Bourgain 等人的证明需假定 $d = 2$. 如何将其证明扩展到一般的整数 d, 仍然是一个挑战性问题.

3.3 结构随机矩阵

由于 Gaussian 矩阵与 Bernoulli 矩阵随机性较强, 确定性矩阵难以证明具有阶数较好的 RIP 性质. 本节中, 我们将介绍介于确定与随机矩阵之间的一种矩阵：结构随机矩阵. 与确定性矩阵相比, 结构随机矩阵多了些随机性, 因而可以证明其具有较好的 RIP 性质, 同时, 结构随机矩阵的随机性较弱, 一般仅具有行随机. 更为重要的是, 在很多实际应用中, 观测矩阵为一结构随机矩阵. 我们在此介绍部分随机 Fourier 矩阵. 我们假定 Ψ 为 $N \times N$ 的离散 Fourier 矩阵. 也就是, Ψ 中的元素为

$$\Psi_{j,k} = \frac{1}{\sqrt{N}}\exp\left(-\frac{2\pi ijk}{N}\right), \qquad j,k \in \{0,\ldots,N-1\}.$$

我们可在矩阵 Ψ 中随机选择 n 行, 得到一个 $n\times N$的矩阵 Ψ_n, 我们称之为部分随机 Fourier 矩阵. 部分随机 Fourier 矩阵具有较强的应用背景. 例如, 很多时候人们观测到的是部分频率信息. 这时候, 观测矩阵就是一部分随机 Fourier 矩阵. 那么, 文 [10] 中作者证明, 矩阵 Ψ_n 高概率的满足 $s=\mathcal{O}(n/(\log N)^6)$ 阶 RIP 性质. 在 [38] 中, 这个结果被改进为 $s=\mathcal{O}(n/(\log N)^4)$. 然而, 人们相信这个结果并非最优的. 因而, 证明矩阵 Ψ_n 高概率的满足 $s=\mathcal{O}(n/\log(N/s))$ 阶 RIP 性质, 仍然是一挑战性问题. 更多的关于结构随机矩阵的介绍, 可参考 [36].

4 宽度与最优编码、解码

假定 $K\subset\mathbb{R}^N$ 是我们感兴趣的信号集合. 我们用 $\mathcal{A}_{n,N}$ 表示所有尺寸为 $n\times N$ 的编码、解码对集合. 前面我们已经介绍了一对具体的编码、解码, 即矩阵 Φ 为 RIP 矩阵, 解码为 Δ_1. 而且我们也看到, 该编码、解码对具有优良的性能. 我们现在考虑如下问题: **对于信号集合 K, 最优编码、解码对 $(\Phi,\Delta)\in\mathcal{A}_{n,N}$ 的性能是什么?** 我们可用严格的数学语言将该问题描述如下: 给定范数 X, $E_n(K)_X$ 是什么? 这里,

$$E_n(K)_X:=\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}}\sup_{x\in K}\|\Delta\Phi x-x\|_X.$$

我们将看到, $E_n(K)_X$ 与 Gelfand 宽度紧密相关. 所谓集合 $K\subset\mathbb{R}^N$ 在附范空间 $(\mathbb{R}^N,\|\cdot\|_X)$ 中 n 阶 Gelfand 宽度, 即指

$$d^n(K)_X:=\inf_{A\in\mathbb{R}^{n\times N}}\sup_{v\in K\cap\ker A}\|v\|_X.$$

下面的定理显示了 $E_n(K)_X$ 与 $d^n(K)_X$ 之间的关联. 其证明可参考 [12].

定理 4.1 *假定 K 是 $\mathbb{R}^N$ 的一个子集且满足 $K=-K$ 及 $K+K=C_0K$, 这里 C_0 是一个大于 0 的常数, 且 $\|\cdot\|_X$ 为一范数. 那么*

$$d^n(K)_X\le E_n(K)_X\le C_0d^n(K)_X,\qquad 1\le n\le N.$$

上面定理显示, 我们可用集合 K 的 Gelfand 宽度的结果来刻画最优编码、解码对的性能. 而 Gelfand 宽度在经典的逼近论中已有较为丰富的研究, 可参考 [34, 35]. 我们下面利用该定理, 导出一个令人感兴趣的结果. 令

$$B_1^N:=\{x\in\mathbb{R}^N:\|x\|_1\le 1\}.$$

一个经典的不等式是:

$$\sigma_s(x)_2\le\frac{1}{\sqrt{s}}\|x\|_1.$$

我们看到, 如果我们选择 $x \in B_1^N$, 那么

$$\sigma_s(x)_2 \le \frac{1}{\sqrt{s}}.$$

通过这个不等式, 我们可有

$$\frac{1}{2\sqrt{s}} \le \sigma_s(B_1^N)_2 \le \frac{1}{\sqrt{s}}. \tag{5}$$

人们已经对 B_1^N 的 Gelfand 宽度进行了深入研究 (参考[22]). 特别是, 存在常数 C_1, C_2 使得

$$C_1 \min\{1, \sqrt{\frac{\log(N/n)}{n}}\} \le d^n(B_1^N)_2 \le C_2 \min\{1, \sqrt{\frac{\log(N/n)}{n}}\}. \tag{6}$$

如果我们希望存在一个常数 C_3, 使得

$$E_n(B_1^N)_2 \le C_3 \sigma_s(B_1^N)_2,$$

那么, 根据定理 4.1, (5) 和 (6), 则有

$$s \le c_0 \frac{n}{\log(N/n)},$$

这里, c_0 为一绝对常数.

注 4.1 *定理 4.1 显示了编码、解码对 (Φ, Δ) 的最优性能与宽度之间的关联. 基于这个关联, 人们能更好的理解压缩感知中编码、解码对的性能. 此外, 用压缩感知中发展的方法, 文 [22] 亦解决了宽度理论中的一些经典问题.*

5 个例最优性

如前所述, 对于 s-稀疏信号 x, 我们一般希望寻找一编码、解码对 $(\Phi, \Delta) \in \mathcal{A}_{n,N}$ 使得 $\Delta(\Phi x) = x$. 那么, 对于一般的信号 $x \in \mathbb{R}^N$, 我们应该设置什么样的恢复误差才比较合理? 一个选择是最佳 s 项逼近误差的常数倍, 也就是 $C_0 \sigma_s(x)_X$, 这里 C_0 是一个绝对常数. 容易看到, 如果 x 为 s-稀疏信号, 那么 $\sigma_s(x)_X = 0$. 本节里, 我们将讨论, 如果选择恢复误差为 $C_0 \sigma_s(x)_X$, 最小观测次数 n 至少为多少?

我们说 (Φ, Δ) 在范数 X 下, 满足 **s-阶个例最优性**, 倘若

$$\|\Delta(\Phi x) - x\|_X \le C_0 \sigma_s(x)_X, \qquad \forall\, x \in \mathbb{R}^N. \tag{7}$$

我们通常选择范数 X 为 ℓ_p 范数. 我们主要考虑编码矩阵为 RIP 矩阵, 解码为 Δ_1 的情形. 我们首先考虑 X 为 ℓ_1 范数.

定理 5.1 ([12]) *令 Φ 是 $3s$-阶 RIP 矩阵, 且 RIP 常数为 $\delta_{3s} \le \delta < (\sqrt{2}-1)^2/3$. 那么,*

$$\|\Delta_1(\Phi x) - x\|_1 \le C_0\sigma_s(x)_1, \qquad \forall x \in \mathbb{R}^N,$$

这里 $C_0 = \frac{2\sqrt{2}+2-(2\sqrt{2}-2)\delta}{\sqrt{2}-1-(\sqrt{2}+1)\delta}$.

从上述定理可看出, 定理 5.1 中定义的编码、解码对在范数 ℓ_1 下具有个例最优性. 如前所述, 我们可以构 $n \times N$ 造矩阵 Φ 满足 s-阶 RIP 条件且有 $n \ge cs\log(N/s)$, 这里 c 是一个固定常数. 因而, 我们只需做 $\mathcal{O}(s\log(N/s))$ 次观测, 就可以得到 ℓ_1 范数下的个例最优性. 那么, 在 $\mathcal{O}(s\log(N/s))$ 次观测的条件下, 我们能够达到 ℓ_2 范数下的个例最优性吗? 文 [12] 表明, 在范数 ℓ_2 下, 即使要达到 1-阶个例最优性, 观测次数 $n \ge N/C_0^2$. 这个结论表明, 如果我们希望在 ℓ_2 范数下达到个例最优性, 那么观测次数与信号的真实维数基本一致. 也就是说, 在 ℓ_2 范数个例最优性的评判标准下, 人们难以在观测次数上“偷工减料”.

注 5.1 *给定一 $x_0 \in \mathbb{R}^N$. 文 [12] 研究了概率意义下的 ℓ_2 个例最优性. 特别的, 文 [12] 证明了, 如果选择 $\Phi \in \mathbb{R}^{n\times N}$ 为 Gaussian 矩阵或者 Bernoulli 矩阵, 这里 $n = \mathcal{O}(s\log(N/n))$. 那么, 存在一解码 Δ, 使得*

$$\|\Delta(\Phi x_0) - x_0\|_2 \le C_0\sigma_s(x_0)_2$$

高概率成立. 更进一步, 文 [44] 证明了, 如果 Φ 为 Gaussian 随机矩阵, 解码 Δ 选择为 P_1 的解, 那么

$$\|\Delta(\Phi x_0) - x_0\|_2 \le C_0\sigma_s(x_0)_2$$

高概率成立.

定理 2.4 给出了如下结果

$$\|\Delta(\Phi x) - x\|_2 \le C_0\frac{\sigma_s(x)_1}{\sqrt{s}}, \qquad \text{对所有} \quad x \in \mathbb{R}^N.$$

这里的编码、解码对为定理 2.4 中的编码、解码对. 注意到, 该结果中左右两边采用了不同的范数. 这启发人们将个例最优性的定义推广到一般范数情形. 我们说(Φ, Δ) 满足 s-阶 (q, p) 个例最优性, 如果

$$\|\Delta(\Phi x) - x\|_p \le C_0\frac{\sigma_s(x)_q}{s^{1/q-1/p}}, \qquad \text{对所有} \quad x \in \mathbb{R}^N.$$

那么, 下面定理给出了一组满足 (q, p) 个例最优性的编码、解码对.

定理 5.2 ([12]) 令 Φ 是满足 $2k+\tilde{k}$-阶 *RIP* 条件, 且 *RIP* 常数 $\delta_{2k+\tilde{k}} \le \delta < 1$, 这里

$$\tilde{k} := k\left(\frac{N}{k}\right)^{2-2/q}.$$

解码 Δ 定义为

$$\Delta(y) := \mathrm{Argmin}_{z\in\ker\Phi}\sigma_s(z)_p.$$

那么, (Φ,Δ) 满足常数为 C_0 的 (p,q) 个例最优性, 这里

$$C_0 = 2^{\frac{1}{p}+\frac{3}{2}}\frac{1+\delta}{1-\delta} + 2^{1+\frac{1}{p}-\frac{1}{q}}.$$

6 算法

本节里, 我们介绍解码实现的一些算法. 如前所述, 我们可将 P_1 转化为线性规划问题. 但是, 由于我们需要求解的问题规模较大, 一些常规的求解线性规划的方法, 如单纯形算法及内点算法等, 并不能达到令人满意的效果. 考虑到问题本身的特殊性, 即矩阵 A 是稠密的, 然而需要恢复的信号 x_0 则是稀疏的. 人们由此构造了一些迭代算法, 如 Bregman 迭代算法 [3, 32, 48], ADM 算法 [47] 及 Proximity 算法 [27, 28] 等. Bregman 迭代算法已经被证明等价于增广的 Lagrangian 方法. 本文主要介绍另外一种解码算法: 贪婪算法. 一般而言, 当矩阵行数远小于列数, 贪婪算法在实际计算中通常有更好的表现.

贪婪算法通常为寻找如下问题的近似解

$$\begin{array}{ll}\min\limits_{x\in\mathbb{R}^N} & \|x\|_0 \\ \text{s.t.} & \|\Phi x - y\|_2 \le \varepsilon.\end{array}$$

其基本思路就是在 Φ 中选择最少的列, 以使其形成对 y 的近似表示. 最常用的一种贪婪算法为 OMP 算法 [26]. 该算法首先计算 y 在当前已选择列张成空间正交投影补, 然后计算该正交投影补与 Φ 中列内积绝对值的大小. 我们通常每次选取使内积绝对值达到最大的列. OMP 算法也有多种变形, 可参考 [37].

我们在 Algorithm 1 中详细描述了 OMP 算法.

我们用 $\mathrm{OMP}_s(y)$ 表示 OMP 算法迭代 s 步所产生的结果. 自然的, 人们关心 OMP 算法的性能 [46, 29, 49]. 鉴于 RIP 矩阵是压缩感知中较为流行的一类矩阵, 人们自然考虑如果选择编码矩阵为 RIP 矩阵, 解码 OMP 算法的性能如何? 因为定理 2.4 是对 ℓ_1 解码性能较好的刻画. 人们希望将定理 2.4 扩展到 OMP 算法. 而这最终在文 [46] 中完成.

Algorithm 1 $\mathrm{OMP}_s(y)$

输入: 编码矩阵 Φ, 向量 y, 最大稀疏度 s

输出: 恢复的信号 x^*.

初始值: $r^0 = y, c^0 = 0, \Lambda^0 = \emptyset, \ell = 0.$

while $\ell \le s$ **do**

  match: $h^\ell = \Phi^T r^\ell$

  identity: $\Lambda^{\ell+1} = \Lambda^\ell \cup \{\mathrm{argmax}_j |h^\ell(j)|\}$

  update: $c^{\ell+1} = \underset{z:\mathrm{supp}(z)\subset\Lambda^{\ell+1}}{\mathrm{argmin}} \|y - \Phi z\|_2$

  $r^{\ell+1} = y - \Phi c^{\ell+1}$

  $\ell = \ell + 1$

end while

$x^* = c^{s+1}$

定理 6.1 ([46]) 假定 $0 < \delta < 1$, 且 Φ 满足 *RIP* 条件 $\delta_{2s} + (1+\delta)\delta_{2\alpha s} \le \delta$. 那么, 对任意的 $x \in \mathbb{R}^N$,

$$\|\mathrm{OMP}_{2(\alpha-1)s}(\Phi x) - x\|_2 \le C_2 \sigma_s(x)_1/\sqrt{s},$$

这里 $\alpha = \lceil 16 + 15\delta \rceil$ 且 $C_2 = 2(1+\delta)(\sqrt{11+20\delta}+1)+3$.

注 6.1 根据定理 *6.1*, 为了使 *OMP* 算法达到 s-阶 $(2,1)$ 个例最优性, *OMP* 算法需要进行约 $50s$ 次迭代 (如果我们选择 δ 接近 *1*). 那么, 一个令人感兴趣的公开问题是, 什么是最小的常数 n_0 , 使得 *OMP* 算法在迭代 $n_0 s$ 次后具有 s-阶 $(2,1)$ 个例最优性? 此外, 文 *[46]* 也考虑了 *OMP* 算法的 (p,q) 个例最优性.

致谢: 本文在袁亚湘院士建议及鼓励下完成, 在此表示感谢.

参考文献

[1] R.G. Baraniuk, M. Davenport, R.A. DeVore and M. Wakin, A simple proof of the restricted isometry property for random matrices. Constr Approx, 2008, 28: 253 - 263.

[2] J. Bourgain, S. J. Dilworth, K. Ford, S. Konyagin and D. Kutzarova, Explicit constructions of RIP matrices and related problems. Duke Math. J, 2011, 159: 145-185.

[3] J. Cai, S. Osher, and Z. Shen, Convergence of the linearized Bregman iteration for ℓ_1-norm minimization. Mathematics of Computation, 2009, 78: 2127-2136.

[4] T. Cai, L. Wang, and G. Xu, Shifting inequality and recovery of sparse signals. IEEE Trans. Signal Process, 2010, 58: 1300 – 1308.

[5] T. Cai, L. Wang, and G. Xu, New Bounds for Restricted Isometry Constants, IEEE Transactions on Information Theory, 2010, 56: 4388 - 4394.

[6] E. Candès, The restricted isometry property and its implications for compressed sensing, C. R. Math. Acad. Sci. Paris, Series I, 2008, 346: 589-592.

[7] E. J. Candès, J. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math., 59(8)(2006)1207-1223.

[8] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory, 2006, 52: 489-509.

[9] E. Candès, T. Tao, Decoding by linear programming, Issue Date: Dec. 2005, 51: 4203 - 4215.

[10] E. J. Candès and T. Tao, Near-optimal signal recovery from random projections and universal encoding strategies, This paper appears in: Information Theory, IEEE Transactions on Issue Date: Dec. 2006, 52: 5406-5425.

[11] E. J. Candés, Y. C. Eldar, D. Needell, and P. Randall, Compressed Sensing with Coherent and Redundant Dictionaries, Applied and Computational Harmonic Analysis, 2011, 31: 59-73.

[12] A. Cohen, W. Dahmen and R. DeVore, Compressed sensing and best k-term approximation, J. Amer. Math. Soc. 2009, 22: 211 – 231.

[13] Chartrand, R., Staneva, V.: Restricted isometry porperties and nonconvex compressive sensing. Inverse Problems 2009, 24: 1-14.

[14] G. Davis, S. Mallat and M. Avellaneda, Adaptive greedy approximations, Constr. Approx., 1997, 13: 57 – 98.

[15] R. DeVore, Deterministic constructions of compressed sensing matrices, Journal of Complexity 2007, 23: 918-925.

[16] D.L. Donoho, “Neighborly polytopes and sparse solutions of underdetermined linear equations, ” Technical report, Department of Statistics, Stanford University, 2005.

[17] M. E. Davies and R. Gribonval, Restricted isometry constants where ℓ_p sparse recovery can fail for $0 < p \leq 1$, IEEE Trans. Inf. Theory, 2009, 55: 2203-2214.

[18] D.L. Donoho and X. Huo, Uncertainty principles and ideal atomic decompositions, IEEE Trans. Inform. Theory , 2011, 47: 2845 – 2862.

[19] M. Elad and A.M. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of bases, IEEE Trans. Inform. Theory, 2002, 48: 2558 – 2567.

[20] H. G. Feichtinger and T. Strohmer, editors. Gabor Analysis and algorithms: Theory and Applications. Birkhäuser, Boston, 1998.

[21] S. Foucart and M. Lai, Sparsest solutions of underdetermined linear systems via ℓ_1-minimization for $0 < q \leq 1$, Appl. Comput. Harmon. Anal., 2009, 26/3: 395-407.

[22] S. Foucart, A. Pajor, H. Rauhut and T. Ullrich, The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$, Jounal of Complexity, 2010, 26: 629-640.

[23] R. Gribonval and M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory 2003, 49: 3320 – 3325.

[24] B. S. Kashin, Widths of certain finite-dimensional sets and classes of smooth functions, Izv. Akad. Nauk SSSR, Ser. Mat. 1977, 41: 334-351; English transl. in Math. USSR Izv. 1978, 11: 317-333.

[25] B. S. Kashin, On widths of octahedron, Uspekhi Matem. Nauk 1975, 30: 251-252 (Russian).

[26] S. Mallat and Z. Zhang. Matching Pursuits with time-frequency dictionaries. IEEE Trans. Signal Process., 1993, 41:3397-3415/

[27] C. A. Micchelli, Lixin Shen, and Yuesheng Xu, Proximity algorithms for image models: denoising, Inverse Problems, 2011, 27.

[28] C. A. Micchelli, Lixin Shen, Yuesheng Xu and Xueying Zeng, Proximity algorithms for the L1/TV image denoising model, Adv. Comput. Math., 2011.

[29] Q. Mo, Y. Shen, Remarks on the Restricted Isometry Property in Orthogonal Matching Pursuit algorithm, arXiv:1101.4458.

[30] Q. Mo, S. Li, New bounds on the restricted isometry constant δ_{2k}, Appl. Comp. Harm. Anal., 2011, 31: 460-468.

[31] B. K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 1995, 24: 227 – 234.

[32] S. Osher, Y. Mao, B. Dong, and W. Yin, Fast linearized Bregman iteration for compressed sensing and sparse denoising, Commun. Math. Sci. 2010, 8: 93-111.

[33] A. Pinkus, On L_1-Approximation, Cambridge Tracts in Mathematics 93, Cambridge University Press, Cambridge, 1989.

[34] A. Pinkus, N-Widths in Approximation Theory, Springer-Verlag, Berlin, 1985.

[35] A. Pinkus, N-widths and Optimal Recovery, Lect. Notes AMS Short Course ed., in: Proc. Symp. Appl. Math., 1986, 36: 51 – 66.

[36] H. Rauhut, Compressive Sensing and Structured Random Matrices, In M. Fornasier, editor, Theoretical Foundations and Numerical Methods for Sparse Recovery, Volume 9 of Radon Series Comp. Appl. Math., pages 1-92. deGruyter, 2010.

[37] L. Rebollo-Neira and Z. Xu, Adaptive non-uniform B-spline dictionaries on a compact interval, Signal Processing，2010, 90.

[38] M. Rudelson and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Communications on Pure and Applied Mathematics, 2008, 61: 1025 – 1045.

[39] Y. Shen, S. Li, Restricted p-isometry property and its application for nonconvex compressive sensing, Adv Comput Math. DOI 10.1007/s10444-011-9219-y.

[40] T. Strohmer and R. Heath, Grassmannian frames with applications to coding and communication, Appl. Comput. Harmon. Anal. 2003, 14: 257-275.

[41] S. J. Szarek. Condition numbers of random matrices. J. Complexity, 1991, 7:131 – 149.

[42] A. Weil, On some exponential sums, PNAS, USA, 1948, 34: 204-207.

[43] L. R. Welch, Lower bounds on the maximum cross-correlation of signals, IEEE Trans. Info. Theory, 1974, 20: 397-399.

[44] P. Wojtaszczyk, Stability and instance optimality for gaussian measurements in compressed sensing, Found. Comput. Math. 2010, 10: 1-13.

[45] Z. Xu, Deterministic sampling of sparse trigonometric polynomials, Journal of Complexity 2011, 27: 133-140.

[46] Z. Xu, A remark about orthogonal matching pursuit algorithm, arXiv:1005.3093.

[47] J. Yang and Y. Zhang, Alternating direction algorithms for ℓ_1-problems in compressed sensing, SIAM J. SCI. COMPUT., Vol. 33, No. 1, pp. 250 – 278.

[48] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, SIAM Journal on Imaging Sciences, 2008, 1: 143-168.

[49] T. Zhang, Sparse recovery with orthogonal matching pursuit under RIP, IEEE Transactions on Information Theory, 2011, 57: 6215-622.