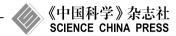
www.springerlink.com

math.scichina.com



国家自然科学基金专栏 综 述

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,北京 100190 E-mail: xuzq@lsec.cc.ac.cn

收稿日期: 2011-12-28: 接受日期: 2012-06-27

国家自然科学基金(批准号: 11171336)及创新研究群体科学基金(批准号: 11021101)资助项目

摘要 压缩感知是近来国际上热门的研究方向. 其主要思想为: 利用信号稀疏性的特征, 通过尽量少的观测信息恢复信号. 压缩感知在多个应用领域, 如医学成像、图像处理、地质勘探等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 (restricted isometry property) 矩阵的构造、Gelfand 宽度、个例最优性及 OMP (orthogonal matching pursuit) 解码等.

关键词 压缩感知 宽度 线性规划

MSC (2010) 主题分类 65T60, 90C25

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

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

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 (1.1) 的解存在且唯一, 我们须选择 $n \ge N$. 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

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

这里 $||x||_0$ 表示 x 中非 0 元素的数目. 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. 本文通常选择 X 为 ℓ_p 范数 $(p \ge 1)$, 并用下标 p 表示 ℓ_p 范数. 当 x_0 中非 0元素较少的时候, 一种较为自然的解码 $\Delta_0(y)$ 是如下规划问题的解:

$$P_0: \min_{x \in \mathbb{R}^N} \|x\|_0$$
s.t. $\Phi x = y$.

这里, $\|x\|_0$ 表示 x 中非 0 元素的数目. <mark>我们用符号 $\Delta_0(y)$ 表示 P_0 的解.</mark> 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. 一个众所周知的结果是: 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 [1,2]. 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:

$$P_1: \min_{x \in \mathbb{R}^N} \|x\|_1$$

s.t.
$$\Phi x = y$$
.

如所知, P₁ 可转化为如下的线性规划问题:

$$P_2: \min_{t \in \mathbb{R}^N} \quad t_1 + t_2 + \dots + t_N$$
 s.t.
$$\Phi x = y,$$

$$-t_j \leqslant x_j \leqslant t_j, \ j = 1, \dots, N,$$

$$t_j \geqslant 0, \ j = 1, \dots, N.$$

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, **对什么样观测矩阵** Φ , P_1 解与 P_0 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

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

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 $T \subset \{1, ..., N\}$ 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. 类似地, 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 $v \in \text{Ker } \Phi$, 均有

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

直观上, 我们将零空间性质理解为 $Ker \Phi$ 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 选择解码为 Δ_1 . 那么, 对任意的 $x \in \Sigma_s$,

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

当且仅当 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 (参考文献 [3]). 与定理 2.2 一致的形式首先出现在文献 [4] 中. 此外, 文献 [5,6] 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**.

我们首先介绍 RIP 性质的定义 ^[7]. <mark>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</mark>

$$(1 - \delta_s) \|x\|^2 \leqslant \|\Phi x\|^2 \leqslant (1 + \delta_s) \|x\|^2 \tag{2.1}$$

对任意的 $x \in \Sigma_s$ 成立. 实际上, (2.1) 等价于 Gram 矩阵 $\Phi_T^{\mathrm{T}}\Phi_T$ 所有特征值位于区间 $[1 - \delta_s, 1 + \delta_s]$, 这里 $\#T \leq s$. 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 (2.1), 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 [8,9] 假定编码矩阵 Φ 满足 2s 阶 RIP 性质, 且 RIP 常数 $\delta_{2s} \leq \sqrt{2} - 1$. 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 $\delta_{2s} \leq \sqrt{2} - 1$. 我们选择解码为 Δ_1 . 那么对任意的 $x \in \mathbb{R}^N$, 我们有

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

此处 C_0 为一常数.

注 2.2 定理 2.3 与 2.4 首先在文献 [9] 中被证明. 但给出的 RIP 常数较为粗糙. 在文献 [8] 中, Candès 将 RIP 常数改进为 $\delta_{2s} \leq \sqrt{2}-1$. 仍有一些文献考虑改进定理 2.3 中的 RIP 常数 $\sqrt{2}-1$ ^[10,11]. 特别是, Mo 和 Li 将该常数改进为 $\delta_{2s} < 0.4931$ ^[12]. 此外, Davies 和 Gribonval 建构一个例子表明, 如果 $\delta_{2s} \geq \frac{1}{\sqrt{2}}$, 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. 在文献 [13] 中, Cai 等考虑了矩阵满足 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, 在文献 [14] 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 $0 ,人们也考虑了如下定义的 <math>\Delta_p$ 解码:

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

这里 $||x||_p := (\sum_{j=1}^N |x_j|^p)^{1/p}$. 事实上, 当 $0 , <math>|| \cdot ||_p$ 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 $^{[15,16]}$.

注 2.4 本文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, 例如在小波基底或者小波框架表示下是稀疏的 (参考文献 [17,18]). 对于此类情况, 文献 [19] 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? 文献 [20] 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

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

则

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

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 $n = \Omega(s\log(\frac{N}{s}))$ 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 $n = \Omega(s\log(\frac{N}{s}))$, 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 $n = \Omega(s\log(\frac{N}{s}))$? 我们将在下节回答该问题.

3 RIP 矩阵

根据定理 2.3 和 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: Gauss 随机矩阵与 Bernoulli 随机矩阵. 所谓 Gauss 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:

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

即服从期望为 0, 方差为 $\frac{1}{n}$ 的 Gauss 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 $-\frac{1}{\sqrt{n}}$, 也就是

$$\Pr\left(\phi_{i,j} = \frac{1}{\sqrt{n}}\right) = \Pr\left(\phi_{i,j} = -\frac{1}{\sqrt{n}}\right) = \frac{1}{2}.$$

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

$$s \leqslant C_1 n / \log(N/s)$$
,

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 ^[21]. 文献 [22,23] 中也给出了定理 3.1 的证明. 一个比较简单的证明方法是文献 [24] 中所介绍的. 此外, 文献 [24] 也给出了 RIP 性质与 Johnson-Lindenstrauss 引理的关联.
- 注 3.2 定理 3.1 表明 Gauss 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 $\Phi=(a_1,\ldots,a_N)\in\mathbb{C}^{n\times N}$,这里 $n\leqslant N$. 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 [25]:

$$\mathcal{M}(\Phi) \geqslant \sqrt{\frac{N-n}{(n-1)N}}.$$
 (3.1)

当等号成立的时候, 我们称矩阵 Φ 为**最优 Grassman 框架**. 文献 [26] 中指出, 只有当 $N \leq n^2$, 等号才有可能成立 (参考文献 [27]).

下面定理显示了矩阵的列相干性与 RIP 性质之间的关联 [28,29].

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 (参考文献 [28,30,31]). 我们在此介绍作者在文献 [31] 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 [32]:

定理 3.3 [32] 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, $1 \le j \le d$, 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 $n \ge 2q+1$ 为素数, 且 $N=(2q+1)^d$. 那 Δ , 矩阵 Φ 的第 j 行定义为

$$\Phi_{j,\cdot} = \left[\frac{\exp(2\pi i \langle x_j, k \rangle)}{\sqrt{n}} \right]_{k \in [-q,q]^d} \in \mathbb{C}^{(2q+1)^d}, \tag{3.2}$$

这里

$$x_i = [j, j^2, \dots, j^d]/n \pmod{1}.$$

下面的定理刻画了由 (3.2) 所定义的矩阵 Φ 的列相干性 [31].

定理 3.4 给定正整数 q 和 d, 令 $n \ge 2q+1$ 为素数, 且 $N = (2q+1)^d$. 假设 $n \times N$ 矩阵 Φ 由 (3.2) 定义. 那么,

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

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

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

组合定理 3.2 和 3.4, 我们有

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

在文献 [31] 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 (3.2) 所定义的确定性矩阵 Φ 的 $\mathcal{O}(\sqrt{n}/d)$.

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

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

因而

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

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 (3.2) 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 Gauss 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

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

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, 文献 [22] 证明, 矩阵 Ψ_n 高概率的满足 $s = \mathcal{O}(n/(\log N)^6)$ 阶 RIP 性质. 在文献 [33] 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, 可参考文献 [34].

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 $A_{n,N}$ 表示所有尺寸为 $n \times N$ 的编码、解码对组成的集合. 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 $(\Phi, \Delta) \in A_{n,N}$ 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. 其证明可参考文献 [35].

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

$$d^n(K)_X \leqslant E_n(K)_X \leqslant C_0 d^n(K)_X, \qquad 1 \leqslant n \leqslant N.$$

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, 可参考文献 [36,37]. 我们下面利用该定理, 导出一个令人感兴趣的结果. 令

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

一个经典的不等式是, 对任意 $x \in \mathbb{R}^N$,

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

我们看到, 如果选择 $x \in B_1^N$, 那么

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

通过这个不等式, 我们可有

$$\frac{1}{2\sqrt{s}} \leqslant \sigma_s(B_1^N)_2 \leqslant \frac{1}{\sqrt{s}}.\tag{4.1}$$

人们已经对 B_1^N 的 Gelfand 宽度进行了深入研究 (参考文献 [20]). 特别是, 存在常数 C_1, C_2 使得

$$C_1 \min\left\{1, \sqrt{\frac{\log(N/n)}{n}}\right\} \leqslant d^n(B_1^N)_2 \leqslant C_2 \min\left\{1, \sqrt{\frac{\log(N/n)}{n}}\right\}. \tag{4.2}$$

如果我们希望存在一个常数 C_3 , 使得

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

那么, 根据定理 4.1, (4.1) 和 (4.2), 则有

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

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, 文献 [20] 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

$$\|\Delta(\Phi x) - x\|_X \leqslant C_0 \sigma_s(x)_X, \qquad \forall \ x \in \mathbb{R}^N.$$
(5.1)

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

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

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, 我们可以构造 $n \times N$ 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? 文献 [35] 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 $x_0 \in \mathbb{R}^N$. 文献 [35] 研究了概率意义下的 ℓ_2 个例最优性. 特别地, 文献 [35] 证明了, 如果选择 $\Phi \in \mathbb{R}^{n \times N}$ 为 Gauss 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

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

高概率成立. 更进一步, 文献 [38] 证明了, 如果 Φ 为 Gauss 随机矩阵, 解码 Δ 选择为 P_1 的解, 那么

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

高概率成立.

定理 2.4 给出了如下结果

$$\|\Delta(\Phi x) - x\|_2 \leqslant C_0 \frac{\sigma_s(x)_1}{\sqrt{s}}, \quad \text{ π in f $x \in \mathbb{R}^N$.}$$

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

$$\|\Delta(\Phi x) - x\|_p \leqslant C_0 \frac{\sigma_s(x)_q}{s^{1/q - 1/p}}, \quad \forall \text{ if } f x \in \mathbb{R}^N.$$

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

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

解码 Δ 定义为

$$\Delta(y) := \underset{z \in \ker \Phi}{\operatorname{Argmin}} \, \sigma_s(z)_p.$$

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 $^{[39-41]}$ 及 ADM (alternating direction method) 算法 $^{[42]}$ 等. Bregman 迭代算法已经被证明等价于增广的 Lagrange 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

$$\min_{x \in \mathbb{R}^N} \quad ||x||_0$$
s.t.
$$||\Phi x - y||_2 \leqslant \varepsilon.$$

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 $^{[43]}$. 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, 可参考文献 $^{[44]}$. 我们在算法 1 中详细描述了 OMP 算法.

我们用 $OMP_s(y)$ 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 $^{[45-47]}$. 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. 而这最终在文献 [45] 中完成.

定理 6.1 [45] 假定 $0 < \delta < 1$,且 Φ 满足 RIP 条件 $\delta_{2s} + (1 + \delta)\delta_{2\alpha s} \leq \delta$. 那么, 对任意的 $x \in \mathbb{R}^N$.

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

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

算法 1 $OMP_s(y)$

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 $\ell \leqslant s$ match: $h^\ell = \Phi^T r^\ell$ identity: $\Lambda^{\ell+1} = \Lambda^\ell \cup \{\operatorname{argmax}_j | h^\ell(j) | \}$ update: $c^{\ell+1} = \underset{z:\operatorname{supp}(z) \subset \Lambda^{\ell+1}}{\operatorname{argmin}} \|y - \Phi z\|_2$ $r^{\ell+1} = y - \Phi c^{\ell+1}$ $\ell = \ell + 1$ end while

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

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_0s 次后具有 s- 阶 (2,1) 个例最优性? 此外, 文献 [45] 也考虑了 OMP 算法的 (p,q) 个例最优性.

7 结论

本文综述了压缩感知基本结果,并介绍了一些新的研究进展. 压缩感知是一个理论及应用皆具潜力的研究方向. 理论方面, 其将逼近论、随机矩阵及离散几何等领域有机的结合在一起. 此外, 压缩感知已经应用于医学成像、断层扫描等, 并有更多的应用领域待探索. 在该方向, 亦有更多的问题值得深入研究, 如确定性 RIP 矩阵的构造 (参考 [28,31])、数据量化对压缩感知的影响 (参考 [48,49])、OMP算法的理论分析 (参考 [45])等. 我们相信对该方向的深入研究将会促进多个理论及应用学科的发展.

致谢 本文在袁亚湘院士建议及鼓励下完成,在此表示感谢.浙江大学李松教授对本文撰写提供了有益的建议,亦表示感谢.

参考文献.

- 1 Davis G, Mallat S, Avellaneda M. Adaptive greedy approximations. Constr Approx, 1997, 13: 57–98
- 2 Natarajan B K. Sparse approximate solutions to linear systems. SIAM J Comput, 1995, 24: 227-234
- 3 Pinkus A. On L_1 -Approximation. Cambridge Tracts in Mathematics 93. Cambridge: Cambridge University Press, 1989
- 4 Gribonval R, Nielsen M. Sparse representations in unions of bases. IEEE Trans Inform Theory, 2003, 49: 3320-3325
- 5 Donoho D L, Huo X. Uncertainty principles and ideal atomic decompositions. IEEE Trans Inform Theory, 2011, 47: 2845-2862

- 6 Elad M, Bruckstein A M. A generalized uncertainty principle and sparse representation in pairs of bases. IEEE Trans Inform Theory, 2002, 48: 2558—2567
- 7 Candès E, Tao T. Decoding by linear programming. IEEE Trans Inform Theory, 2005, 51: 4203-4215
- 8 Candès E. The restricted isometry property and its implications for compressed sensing. C R Math Acad Sci Paris Ser I. 2008, 346: 589–592
- 9 Candès E J, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Comm Pure Appl Math, 2006, 59: 1207–1223
- 10 Cai T, Wang L, Xu G. Shifting inequality and recovery of sparse signals. IEEE Trans Signal Process, 2010, 58: 1300-1308
- 11 Foucart S, Lai M. Sparsest solutions of underdetermined linear systems via ℓ_1 -minimization for $0 < q \le 1$. Appl Comput Harmon Anal, 2009, 26: 395–407
- 12 Mo Q, Li S. New bounds on the restricted isometry constant δ_{2k} . Appl Comp Harm Anal, 2011, 31: 460–468
- 13 Cai T, Wang L, Xu G. New bounds for restricted isometry constants. IEEE Trans Inform Theory, 2010, 56: 4388-4394
- 14 Donoho D L. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report, Department of Statistics, Stanford University, 2005
- 15 Shen Y, Li S. Restricted *p*-isometry property and its application for nonconvex compressive sensing. Adv Comput Math, doi: 10.1007/s10444-011-9219-y
- 16 Chartrand R, Staneva V. Restricted isometry properties and nonconvex compressive sensing. Inverse Problems, 2009, 24: 1–14
- 17 Shen Z. Wavelet frames and image restorations. In: Bhatia R, ed. Proceedings of the International congress of Mathematicians, vol. IV. Hyderabad: Hindustan book Agency, 2010, 2834–2863
- 18 Shen Z W, Xu Z Q. On B-spline framelets derived from the unitary extension principle. ArXiv:1112.5771
- 19 Candés E J, Eldar Y C, Needell D, et al. Compressed sensing with coherent and redundant dictionaries. Appl Comput Harmonic Anal, 2011, 31: 59–73
- 20 Foucart S, Pajor A, Rauhut H, et al. The Gelfand widths of ℓ_p -balls for 0 . J Complexity. 2010, 26: 629–640.
- 21 Kashin B S. 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
- Candès E J, Tao T. Near-optimal signal recovery from random projections and universal encoding strategies. IEEE Trans Inform Theory, 2006, 52: 5406-5425
- 23 Szarek S J. Condition numbers of random matrices. J Complexity, 1991, 7: 131–149
- 24 Baraniuk R G, Davenport M, DeVore R A, et al. A simple proof of the restricted isometry property for random matrices. Constr Approx, 2008, 28: 253–263
- 25 Welch L R. Lower bounds on the maximum cross-correlation of signals. IEEE Trans Inform Theory, 1974, 20: 397–399
- 26 Feichtinger H G, Strohmer T, eds. Gabor Analysis and Algorithms: Theory and Applications. Boston: Birkhäuser, 1998
- 27 Strohmer T, Heath R. Grassmannian frames with applications to coding and communication. Appl Comput Harmon Anal, 2003, 14: 257–275
- 28 DeVore R. Deterministic constructions of compressed sensing matrices. J Complexity, 2007, 23: 918-925
- 29 Bourgain J, Dilworth S J, Ford K, et al. Explicit constructions of RIP matrices and related problems. Duke Math J, 2011, 159: 145–185
- 30 Kashin B S. On widths of octahedron (in Russian). Uspekhi Matem Nauk, 1975, 30: 251–252
- 31 Xu Z. Deterministic sampling of sparse trigonometric polynomials. J Complexity, 2011, 27: 133-140
- 32 Weil A. On some exponential sums. Proc Natl Acad Sci USA, 1948, 34: 204-207
- 33 Rudelson M, Vershynin R. On sparse reconstruction from Fourier and Gaussian measurements. Comm Pure Appl Math, 2008, 61: 1025–1045
- 34 Rauhut H. Compressive sensing and structured random matrices. In: Fornasier M, ed. Theoretical Foundations and Numerical Methods for Sparse Recovery, vol. 9 of Radon Series Comp Appl Math. Berlin: de Gruyter, 2010, 1–92
- 35 Cohen A, Dahmen W, DeVore R. Compressed sensing and best k-term approximation. J Amer Math Soc, 2009, 22: 211-231
- $36\,\,$ Pinkus A. n-Widths in Approximation Theory. Berlin: Springer-Verlag, $1985\,$
- $37\,\,$ Pinkus A. n-Widths and optimal recovery. Proc Sympos Appl Math, 1986, $36:\,51\textsc{-}66$
- 38 Wojtaszczyk P. Stability and instance optimality for Gaussian measurements in compressed sensing. Found Comput Math, 2010, 10: 1–13
- 39 Cai J, Osher S, Shen Z. Convergence of the linearized Bregman iteration for ℓ_1 -norm minimization. Math Comput,

- 2009, 78: 2127-2136
- 40 Osher S, Mao Y, Dong B, et al. Fast linearized Bregman iteration for compressed sensing and sparse denoising. Commun Math Sci, 2010, 8: 93–111
- 41 Yin W, Osher S, Goldfarb D, et al. Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing. SIAM J Imaging Sciences, 2008, 1: 143–168
- 42 Yang J, Zhang Y. Alternating direction algorithms for ℓ_1 -problems in compressed sensing. SIAM J Sci Comput, 2011, 33: 250–278
- 43 Mallat S, Zhang Z. Matching Pursuits with time-frequency dictionaries. IEEE Trans Signal Process, 1993, 41: 3397–3415
- 44 Rebollo-Neira L, Xu Z. Adaptive non-uniform B-spline dictionaries on a compact interval. Signal Processing, 2010, 90: 2308–2313.
- 45 Xu Z. A remark about orthogonal matching pursuit algorithm. ArXiv:1005.3093
- 46 Mo Q, Shen Y. Remarks on the restricted isometry property in orthogonal matching pursuit algorithm. ArXiv:1101.4458
- 47 Zhang T. Sparse recovery with orthogonal matching pursuit under RIP. IEEE Trans Inform Theory, 2011, 57: 6215–622
- 48 Wang Y, Xu Z. The performance of PCM quantization under tight frame representations. SIAM J Math Anal, in press
- 49 Güntürk S, Powell A, Saab R, et al. Sobolev duals for random frames and sigma-delta quantization of compressed sensing measurements. ArXiv:1002.0182v1

Compressed sensing: a survey

XU ZhiQiang

Abstract Compressed sensing is an active research topic recently. It is related with many different mathematical subjects, such as approximation theory, optimization, random matrixes and discrete geometry, etc., and produces many beautiful results. Moreover, it can be used in medical imaging and CT, etc. In this paper, we will introduce the basic results in compressed sensing and also present many new developments. In particular, we will introduce the RIP encoding and ℓ_1 decoding, and the performance of OMP algorithm.

Keywords compressed sensing, width, linear programming

MSC(2010) 65T60, 90C25

doi: 10.1360/012011-1043