

第3章 压缩感知与成像

2006年,Donoho和Candès等提出一种新颖的信号采样和处理理论,即压缩感知(CS)理论[Donoho 2006]、[Candès 2006]。该理论将信号采样和压缩结合进行,提高了信号采集和信息处理能力,突破了奈奎斯特-香农采样定理的瓶颈,大大缓解了数据采集、存储、传输和分析的压力。

压缩感知理论指出,当信号在某个变换域里可压缩或稀疏时,通过采集少量的信号的非自适应线性投影值,构建数学重构模型并利用重构算法求解,可以较精确地重构原信号。简单地说,稀疏的或具有稀疏表达的有限维数的信号可以利用远少于奈奎斯特-香农采样数量的线性、非自适应的测量值而无失真地重建出来。

压缩感知的理论框架和研究内容主要涉及稀疏表达、测量编码和解码重构三方面,即用于稀疏表达(表示)的稀疏编码、用于测量编码的测量矩阵和用于重建信号的重构算法。首先,稀疏表达是压缩感知理论的先验/先决条件。其次,测量编码是压缩感知硬件实现的关键,但这里并非直接测量信号本身,而是经过测量矩阵将信号投影到低维空间,并保持信号原本的结构信息。最后,解码重构是压缩感知理论的核心,指利用尽量少的测量值,快速、鲁棒、精确地重构原信号。

根据上述的讨论,本章各节将安排如下。

3.1节介绍压缩感知的整体流程,并引出3个关键处理步骤。另外,还对压缩感知与奈奎斯特-香农采样在流程和特性方面进行了对比。

3.2节介绍对信号的稀疏表达,通过把信号变换到一个新的基或框架下,使其非零系数的个数远远少于原始信号的项数时,就可以实现压缩感知了。

3.3节讨论测量编码模型,不仅介绍了针对本身稀疏信号的测量模型,还介绍了具有普适性的测量模型。另外,还对测量矩阵应满足的两个特性进行了分析。

3.4节介绍了基本的重构模型,分析了无噪声和有噪声时重构稀疏信号的原理,并列了几类典型的重构算法。

3.5节讨论了曾与压缩感知并行发展,密切相关又互相影响的稀疏编码与字典学习问题。稀疏编码与信号的稀疏表达对应,而字典学习的目标就是要实现对信号的稀疏表达。

3.6节列举两个压缩感知在成像应用方面的典型示例:一个是单像素相机的设计;另一个是利用压缩感知的磁共振成像。

3.1 压缩感知概述

压缩感知的整体流程框图可见图3.1.1,其中第一行的3个方框分别代表压缩感知的3个关键处理步骤(每个方框的左边是输入,右边是输出),第二行的3个方框分别指示实现压

缩感知要完成的 3 个工作(分别对应 3 个关键处理步骤)。

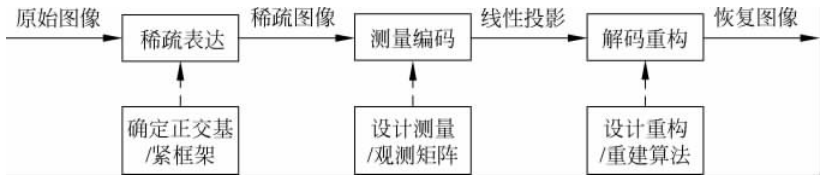


图 3.1.1 压缩感知流程

压缩感知有 3 个关键处理步骤：

(1) 稀疏表达

判断信号是否具有稀疏性是实现压缩感知的前提。事实上，自然信号很少是真正稀疏的。当认为某个信号稀疏时，严格地说是指这个信号可以通过稀疏信号近似表达。这里的工作就是要找到对信号可压缩的某个正交基或紧框架(见下)，从而可实现对信号的稀疏表达。从数学角度来说，是要构建稀疏变换矩阵，实现对多维数据进行线性分解的表达方法。

(2) 测量编码

对信号采用一个与稀疏变换矩阵不相关(与稀疏变换的基不相关且平稳)的线性投影进行稀疏测量/观测，从而获得感知测量值。对感知测量值的表达常称为编码，所以统称测量编码。进行稀疏变换的测量矩阵是压缩感知理论能否成功实现的关键。这里对观测矩阵的设计不仅关系到压缩和采样过程的快慢，而且要保证对稀疏向量降维时重要信息不遭到破坏，从而提高恢复信号的准确性。

(3) 解码重构

这是压缩感知模型求解的保证，也是压缩感知的核心部分。其含义是运用压缩测量的低维数据(感知测量值)精确地重构高维的原始信号。这里的工作是要设计快速的重构算法，从一个非常少的线性观测中恢复信号，重构算法的优劣决定了恢复信号的质量。

例 3.1.1 压缩感知与奈奎斯特-香农采样的对比

从信号处理的角度看，压缩感知与奈奎斯特-香农采样有很多区别。图 3.1.2 同时画出了它们的基本流程，左边为香农采样的流程，高速采样后压缩以减少数据量，在需要原始图像时对压缩结果进行解压；右边为压缩感知采样的流程，借助稀疏变换实现低速采样，利用观测矩阵实现投影降维，在需要原始信息时将图像重构出来。

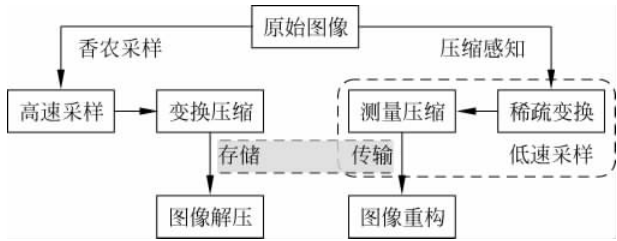


图 3.1.2 压缩感知与奈奎斯特-香农采样的流程对比

表 3.1.1 给出对压缩感知与奈奎斯特-香农采样在 5 个特性方面的对比情况。

表 3.1.1 压缩感知与奈奎斯特-香农采样的特性对比

特 性	奈奎斯特-香农采样	压缩感知采样
必要条件	信号带宽有限：才可能无损恢复	信号具有稀疏性/可压缩性：才可能实现恢复
采样方式	直接采样：通过 Sinc 函数直接与目标信号进行内积来获取采样	非直接采样：通过采样/测量矩阵与目标信号进行乘积间接地获取采样/测量值
采样特点	均匀采样：用目标信号中最高频率 2 倍以上的采样率对目标信号进行等间隔采样	非均匀采样：用随机非均匀的采样方式借助非相关测量矩阵对目标信号进行采样/测量
采样个数	信号频率决定采样个数：采样率至少为目标信号中最高频率的 2 倍，以实现无失真采样	稀疏性决定采样个数：如果目标信号中仅有少量个非零信号，则只需用很少的采样/测量个数以恢复原始信号
从采样恢复信号	所采即所得：只需用 Sinc 函数插值就可直接重建原始信号	需要重建步骤：因没有直接对目标信号采样，所以需要（如利用基于 L_1 范数最小化的步骤）进行信号恢复工作

□

3.2 稀疏表达

对信号的稀疏表达是实现压缩感知的先决条件。需要指出的是，虽然自然信号都有一定的稀疏性，但很少有信号是真正稀疏的，直接满足压缩感知的要求。一般认为某个信号是稀疏的时候，是指这个信号可以通过稀疏信号来近似地表达。下面来具体讨论。

先复习、介绍和讨论一些预备知识。

1. 矢量空间

下面的讨论中将有限域中的离散信号看成分布在 N 维欧氏空间中的矢量，将这个空间记为 $\mathbf{R}^N$ 。在 $\mathbf{R}^N$ 中，范数是一个重要描述。矢量 $\mathbf{x}$ 的范数 L_p ($p \geq 1$) 可定义如下：

$$\|\mathbf{x}\|_p = \begin{cases} \left(\sum_{i=1}^N |x_i|^p \right)^{1/p} & p \in [1, \infty) \\ \max |x_i| & p = \infty \end{cases} \quad i = 1, 2, \dots, N \quad (3.2.1)$$

其中， $|S|$ 为集合 S 的基数，即集合 S 中元素的个数。

常用的几个范数分别为 L_1 , L_2 和 L_∞ 。它们的特性分别如图 3.2.1(a)、(b)、(c) 所示（可对比上册书中图 3.1.5）。在 $p < 1$ 的情况下，式 (3.2.1) 中定义的范数已经无法满足三角不等式，所以它本质上是拟范数。拟范数所对应的 $\mathbf{R}^2$ 中的单位圆不再是凸集。例如，拟范数 $L_{1/2}$ 所对应的 $\mathbf{R}^2$ 中的单位圆 $\{\mathbf{x}: \|\mathbf{x}\|_{1/2} = 1\}$ 如图 3.2.1(d) 所示。又如，拟范数 L_0 所对应的 $\mathbf{R}^2$ 中的单位圆 $\{\mathbf{x}: \|\mathbf{x}\|_0 = 1\}$ 如图 3.2.1(e) 所示。这里，拟范数 $\|\mathbf{x}\|_0 = |\text{supp}(\mathbf{x})|$ ，其中 $\text{supp}(\mathbf{x}) = \{i: x_i \neq 0\}$ 表示 $\mathbf{x}$ 的支撑集。

例 3.2.1 不同范数在描述逼近误差时的不同表现

可以采用范数来描述信号的强度或误差的大小。假设已知一个信号 $\mathbf{x} \in \mathbf{R}^2$ ，希望用一个 1-D 仿射空间 A 中的点来逼近它。如果采用 L_p 范数来衡量这种逼近误差，那么任务就

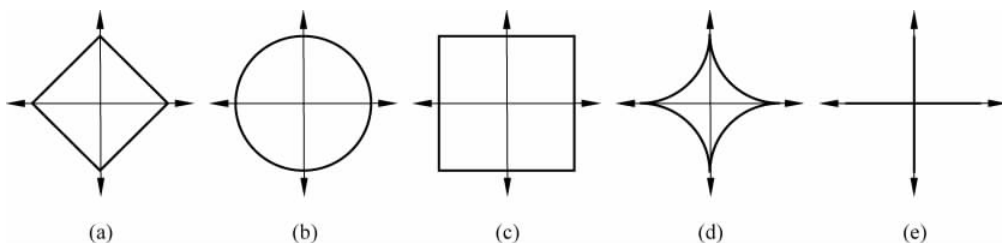


图 3.2.1 不同范数在 $\mathbf{R}^2$ 中的单位圆

是找到 $\mathbf{x}' \in A$ 使得 $\|\mathbf{x} - \mathbf{x}'\|_p$ 最小。这里对参数 p 的选择很关键,不同的 p 值将使得逼近误差具有不同的特性和表现。为了找出在 A 中最接近 $\mathbf{x}$ 的点,可以想象一个以 $\mathbf{x}$ 为中心的 L_p 球不断地膨胀,它碰到 A 的点即为在 L_p 条件下最接近 $\mathbf{x}$ 的点 $\mathbf{x}'$ 。如图 3.2.2 所示,图(a)对应 L_1 范数,图(b)对应 L_2 范数,图(c)对应 L_∞ 范数,图(d)对应 $L_{1/2}$ 拟范数,图(e)对应 L_0 拟范数。

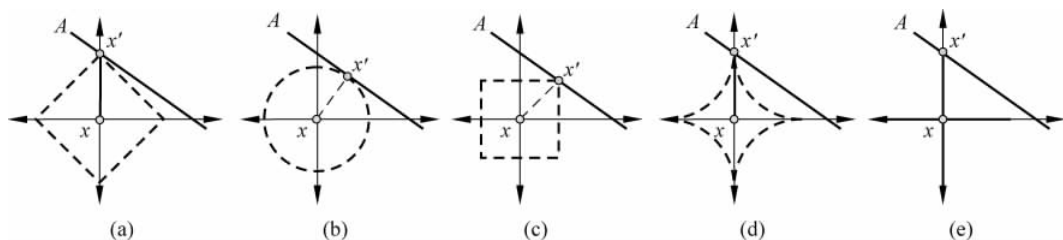


图 3.2.2 4 种 p 值的 L_p 在描述逼近误差时的不同表现

从图 3.2.2 可见,当 p 较大时,误差被均匀地扩散到 2-D 空间中(与 $\mathbf{x}$ 不在同一水平或垂直轴上);当 p 较小时,这种误差的衡量方式将会有很大概率使所选择的 $\mathbf{x}'$ 与 $\mathbf{x}$ 位于同一水平或垂直轴上,即非对称地把误差缩小到 1-D 空间中(减少了一个维度),从而促进稀疏特性的产生。□

2. 基和框架

考虑矢量集 $\Phi = \{\phi_i\}, i=1,2,\dots,N$,如果其中的矢量可以生成矢量空间 $V = \mathbf{R}^N$,而且矢量 ϕ_i 之间是非线性相关的,则 Φ 可以称为有限维矢量空间 V 中的基。换句话说,在矢量空间 V 中的任一个矢量都可以通过 Φ 中矢量的线性组合唯一地表达,而且这种线性表达的系数可以通过使用信号和该矢量空间中基的内积来表示。

表示矢量集的符号也可用来表示矩阵,如 Φ 可表示由列向量 ϕ_i 构成的矩阵,其大小为 $N \times N$ 。标准正交集就是一种特殊的基,定义为: 矢量集 $\Phi = \{\phi_i\}$,其中所有矢量之间是正交的且每个基的范数都为单位 1,即 $\Phi^T \Phi = I$, I 表示 $N \times N$ 的单位矩阵。此时,对于任何属于该矢量空间中的矢量 $\mathbf{x}$,可以很容易地计算出在该标准正交集表示下的系数 $\mathbf{a}$,即 $\mathbf{a} = \Phi^T \mathbf{x}$ 。

把基的概念推广到一些可能线性相关的矢量集就形成列框架,即矢量集 $\Phi = \{\phi_i\}_{i=1}^N$ 且 $\phi_i \in \mathbf{R}^d$,其中 $d < N$,相当于矩阵 $\Phi \in \mathbf{R}^{d \times N}$,对所有矢量 $\mathbf{x} \in \mathbf{R}^d$ 满足

$$L \|\mathbf{x}\|_2^2 \leq \|\Phi^T \mathbf{x}\|_2^2 \leq R \|\mathbf{x}\|_2^2 \quad (3.2.2)$$

其中, $0 < L \leq R < \infty$ 。

注意, $L > 0$ 表示矩阵 Φ 中的行矢量一定是线性独立的。如果 L 被选为使这个不等式成

立的可能存在的最大值, R 被选为使这个不等式成立的可能存在的最小值, 则把它们称为无框架界。如果 $L=R$, 则这个框架称为紧框架。如果 Φ 是一个有限维矩阵, 则 L 和 R 分别对应 $\Phi\Phi^T$ 的最小特征值和最大特征值。由于框架具有一定的冗余性, 所以它可以对目标数据提供更为丰富的表达, 即针对一个目标信号矢量 x , 存在无数个系数矢量 a , 使得 $x=\Phi a$ 。

3. 稀疏性表达

为了更精炼地表达一个信号, 可以把信号变换到一个新的基或框架下, 当非零系数的个数远远少于原始信号的项数时, 可以把这些少量的非零系数称为原始信号的稀疏性表达。

在压缩感知的理论体系中, 稀疏信号模型可以确保高倍的压缩率。只要预先知道目标信号在已知的基或框架下具有稀疏性表达, 就可以无失真地重建原始信号。在稀疏性表达中, 常把基或框架称为字典或过完备字典, 而把其中的矢量元素称为原子。

从数学的角度, 当信号 x 中最多有 K 个非零的值时, 称信号 x 是 K 稀疏的, 即 $\|x\|_0 \leq K$, 可以采用 $\Sigma_K = \{x: \|x\|_0 \leq K\}$ 来表示所有 K 稀疏信号的集合。另外, 对一些本身并不稀疏但在一些基矩阵 Φ 中具有稀疏性表达的信号, 这时只要 $x=\Phi a$, 其中 $\|a\|_0 \leq K$, 则仍可把这些信号看成 K 稀疏的。

3.3 测量矩阵及特性

下面先比较讨论传统方法和压缩感知方法是如何获取信号的, 然后讨论测量矩阵应具有的特性, 最后分析如何构建需要的测量矩阵。

3.3.1 采样/测量模型

根据传统的奈奎斯特-香农采样模型, 对一个只包含 K 个非零值的信号 $x \in \mathbf{R}^N$, 需要采集 N 个值才能保证完全保持 x 中的信息。具体的采样过程可用采样矩阵 Φ 与信号矢量 x 的相乘来表示, 即

$$y = \Phi x \quad (3.3.1)$$

这个采样过程可用图 3.3.1 来表示。此时, 采样矩阵对应一个对角单位矩阵, 它将原始信号逐一映射到采样值矢量 y 。需要注意, 此时在 $K \ll N$ 的情况下, 仍需使用 N 个采样值, 这必然带来很大的浪费。

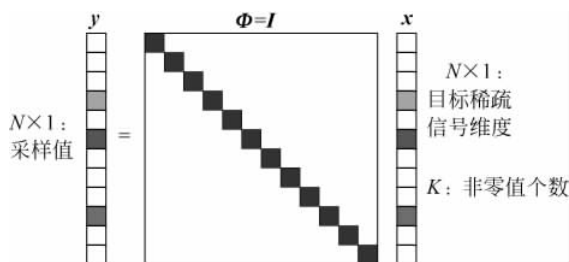


图 3.3.1 奈奎斯特-香农采样模型

压缩感知采样并不对原始信号逐一进行, 而是通过一个测量矩阵 Φ 获取对 x 的 M 个线性观测/测量。这个过程仍可用式 (3.3.1) 的形式表示, 但其中 Φ 是一个固定的 $M \times N$ 矩

阵,采样所得的测量值 $y \in \mathbf{R}^M$ 。先假设信号 x 本身是稀疏的,则此时的测量模型如图 3.3.2 所示,其中 Φ 对应一个降维的投影操作,即把 $\mathbf{R}^N$ 映射到 $\mathbf{R}^M$ 中。一般有 $K < M \ll N$,即矩阵 Φ 的列数远多于行数。

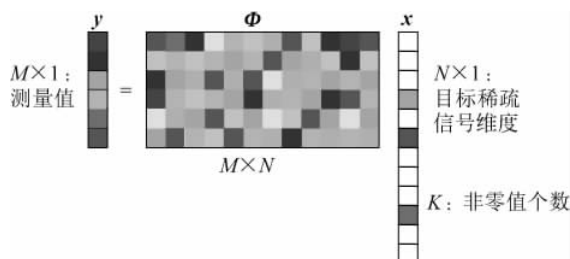


图 3.3.2 针对本身稀疏信号的测量模型

由前面对稀疏表达的讨论可知,对本身并不稀疏的信号,可通过将其变换到一个新的基或框架下而使其表现出稀疏特性(可理解为用变换挖掘出了信号的稀疏性),从而仍可构建压缩感知框架下的测量模型。换句话说,如果 x 本身不是稀疏信号,但在变换域 Ψ 中体现出稀疏性,即 $x = \Psi z$,则将 Φ 和 Ψ 合并,而测量过程仍可用式(3.3.1)的形式来表示。不过这种推广的、具有普适性的测量模型用图 3.3.3 来表示更为清晰。前面针对本身稀疏信号的测量模型对应 $\Psi = I$ 的情况。

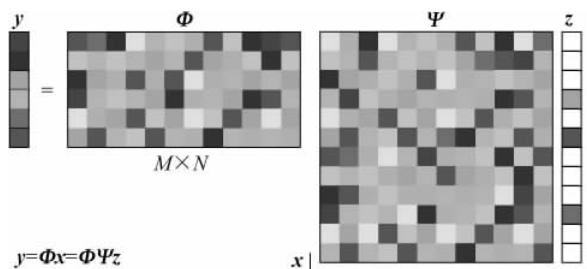


图 3.3.3 具有普适性的测量模型

3.3.2 测量矩阵特性

要实现压缩感知,测量矩阵应满足一定的特性[李 2015]。

1. 零空间及其特性

对一个矩阵 $\Phi: \mathbf{R}^N \rightarrow \mathbf{R}^M$,其零空间可定义为 $\mathcal{N}(\Phi)$

$$\mathcal{N}(\Phi) = \{z: \Phi z = 0\} \quad (3.3.2)$$

对任意稀疏信号 x ,如果希望基于测量值 $y = \Phi x$ 无失真地重建该稀疏信号 x ,则对任意两个矢量 x 和 $x' \in \Sigma_K = \{z: \|z\|_0 \leq K\}$,一定有 $\Phi x = \Phi x'$,否则基于测量值 y 将无法区别 x 和 x' 。假设 $\Phi x = \Phi x'$,则 $\Phi(x - x') = 0$,其中 $x - x' = h \in \Sigma_{2K}$ 。由此可见,用矩阵 Φ 可以唯一表达 x 的充分必要条件是 Φ 的零空间 $\mathcal{N}(\Phi)$ 中不含有任何 Σ_{2K} 中的元素,即 $\mathcal{N}(\Phi)$ 与 Σ_{2K} 的交集是空集。

当要处理的是近似稀疏信号(其含有少数几个明显的非零元素,而其他元素则非常接近零,自然图像经小波变换后的系数就是这样的信号)时,需要对测量矩阵 Φ 的零空间引入一

些更为严格的条件,以确保 $\mathcal{N}(\Phi)$ 中不应包含稀疏的矢量,也不应包含可以压缩的近似稀疏的矢量。

现在假设 $S \subset \{1, 2, \dots, N\}$ 是一个索引集的子集, $S^c = \{1, 2, \dots, N\} \setminus S$ 是其相应的补集, 则矢量 $\mathbf{x}_S$ 表示长度为 N 的矢量, 且这个矢量中所有下标属于集合 S^c 的元素都被设为 0。类似地, 矩阵 Φ_S 表示长度为 $M \times N$ 的矢量, 且其所有下标属于集合 S^c 的列矢量都被设为零矢量。如果存在一个常数 $C > 0$, 使得下式对所有 $\mathbf{h} \in \mathcal{N}(\Phi)$ 和所有 $|S| \leq K$ 的 S 都成立:

$$\|\mathbf{h}_S\| \leq C \frac{\|\mathbf{h}_{S^c}\|_1}{\sqrt{K}} \quad (3.3.3)$$

则称矩阵 Φ 满足 K 阶零空间特性。

现在用 $\Delta: \mathbf{R}^M \rightarrow \mathbf{R}^N$ 表示一种基于信号稀疏性的重建算法, 即基于 M 个测量值重建出原始 N 个目标信号的算法, 而且 $M \ll N$ 。对于所有的 $\mathbf{x}$, 可以要求该算法确保下式成立:

$$\|\Delta(\Phi \mathbf{x}) - \mathbf{x}\|_2 \leq C \frac{\sigma_K(\mathbf{x})_p}{\sqrt{K}} \quad p = 1 \quad (3.3.4)$$

其中, $\sigma_K(\mathbf{x})_p = \min \|\mathbf{x} - \mathbf{x}'\|_p$, 表示去除矢量 $\mathbf{x}$ 的 K 个幅值最大的元素后的 p 范数, 即如果 S 表示 $\mathbf{x}$ 的 K 个最大元素的下标集合, 则 $\sigma_K(\mathbf{x})_p = \|\mathbf{x}_{S^c}\|_p$ 。

式 (3.3.4) 可以保证该重建算法能够正确地重建出所有可能的、有 K 个非零元素的信号, 同时也能使得利用重建的稀疏信号来近似逼近非稀疏信号有一定的鲁棒性, 所以称为具有普适性的重建条件。另外可以证明, 如果 (Φ, Δ) 满足式 (3.3.4), 则矩阵 Φ 满足 $2K$ 阶零空间特性。

2. 约束等距特性

零空间特性是确保重建的必要条件, 但推导时并没有考虑噪声的影响。当测量值有噪声或在量化时引入误差时, 需要讨论更为严格的重建条件。为此, 定义 K 阶约束等距特性 (RIP) 如下。

如果存在 $\delta_K \in (0, 1)$, 使得

$$(1 - \delta_K) \|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta_K) \|\mathbf{x}\|_2^2 \quad (3.3.5)$$

对所有 $\mathbf{x} \in \Sigma_K \{\mathbf{x}: \|\mathbf{x}\|_0 \leq K\}$ 都成立, 则称矩阵 Φ 满足 K 阶约束等距特性, δ_K 称为矩阵 Φ 的约束等距常数 (δ_K 是对所有 K 阶稀疏矢量 $\mathbf{x}$ 均满足的最小整数)。

矩阵的 RIP 是针对两个参数 δ_K 和 K 而言的。如果说某一个矩阵满足 RIP, 即对特定的 δ_K 和 K , 式 (3.3.5) 对任意 $\mathbf{x} \in \Sigma_K$ 均成立。如果矩阵 Φ 满足 $2K$ 阶约束等距特性, 则可以把式 (3.3.5) 解释为: 任意一对 K 阶稀疏的矢量经过矩阵 Φ 的线性变换后, 它们之间的欧氏距离几乎保持不变, 即测量矩阵近似地保持两个 K 阶稀疏矢量间的欧氏距离, 这一特性对克服噪声起重要作用。这同时也表明 K 阶稀疏的矢量不在测量矩阵的零空间中, 因为如果在其中, 则信号就无法重建了。约束等距特性也可以表示成: 随机地从测量矩阵中选取 $2K$ 列, 所构成的子集只能是“近似正交”的。这是因为这个子集的行列长度不同, 所以不可能是真正的正交。

假设利用测量矩阵 Φ 获取一维 K 阶稀疏矢量 $\mathbf{x}$ 的线性测量值 $\mathbf{y}$, 当 $\delta_{2K} < 1$ 时, 就可以基于测量矢量 $\mathbf{y} = \Phi \mathbf{x}$ 重建出原始 K 阶稀疏矢量, 这个重建出的 K 阶稀疏矢量是满足方程 $\mathbf{y} = \Phi \mathbf{x}$ 的唯一稀疏解, 即非零元素个数最少。所以, 如果要重建出独一无二的 K 阶稀疏矢量, 必须要确保 $\delta_{2K} < 1$, 同时约束等距常数 δ_K 不可能为 0, 因为如果 $\delta_{2K} = 0$, 则说明测量矩阵

Φ 是标准正交矩阵,但这是不可能的,因为测量矩阵的列数远大于行数。

约束等距特性指出:所求解的精度和稀疏度是取决于测量矩阵或约束等距常数的。但是,某个测量矩阵是否满足这个特性并无法通过计算的方式来判断,因为这是一个 NP 难题。

如果一个矩阵满足约束等距特性,则它一定满足零空间特性,即约束等距特性比零空间特性更为严格。

如何构建满足 RIP 的测量矩阵呢?固定地构建满足 K 阶 RIP 且尺寸为 $M \times N$ 的测量矩阵是可能的,但要求相对较大的 M 。如果在构建过程中引入随机性,可以降低对大测量个数 M 的要求。这些通过随机方法构建的测量矩阵也常称为通用测量矩阵。可以证明,对由亚高斯随机分布产生的测量矩阵所测得的矢量范数高度集中于原始信号的均值,所以服从亚高斯随机分布的矩阵满足 RIP。使用这样得到的测量矩阵就可以实现利用最少采样个数来完成采样的目的。

3.4 解码重构

重构也称重建,目标是要基于很少的线性观测值来重建出稀疏的目标信号。重构是模型求解的保证。下面先讨论重构的原理,再介绍各种典型的重构算法。

3.4.1 重构原理

先介绍一种直观的重构模型,借此讨论重构的原理以及分析对无噪声稀疏信号重构的问题和对有噪声稀疏信号重构的问题。

1. 基本重构模型

假设矢量 $\mathbf{x} \in \mathbf{R}^N$ 是一个长度为 N 的稀疏信号,测量矩阵 $\Phi: \mathbf{R}^N \rightarrow \mathbf{R}^M$ 已知,现在要基于很少数量的测量值 $\mathbf{y} \in \mathbf{R}^M, M < N$ 重建原始目标信号 $\mathbf{x}$ 。

最直接的重构方法可描述成:

$$\begin{cases} \min \|\mathbf{x}\|_0, & \text{s. t. } \mathbf{y} = \Phi \mathbf{x}, \text{测量值无噪声的情况} \\ \min \|\mathbf{x}\|_0, & \text{s. t. } \|\Phi \mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \text{测量值存在少量有界噪声的情况} \end{cases} \quad (3.4.1)$$

如果原始目标信号在某个正交变换域或稀疏字典 Ψ 中体现出稀疏特性,即 $\mathbf{x} = \Psi \mathbf{z}$,其中 $\|\mathbf{z}\|_0 \leq K$,则可将式(3.4.1)改写为

$$\begin{cases} \min \|\mathbf{z}\|_0, & \text{s. t. } \mathbf{y} = \Phi \Psi \mathbf{z}, \text{测量值无噪声的情况} \\ \min \|\mathbf{z}\|_0, & \text{s. t. } \|\Phi \Psi \mathbf{z} - \mathbf{y}\|_2 \leq \epsilon, \text{测量值存在少量有界噪声的情况} \end{cases} \quad (3.4.2)$$

可见,如果令 $\Theta = \Phi \Psi$,则本质上前面两式是等价的。下面仅考虑 $\Psi = \mathbf{I}$ 的情况。

求解式(3.4.1)或式(3.4.2)是一个组合优化问题,所以是一个 NP 难题,可采用凸的 L_1 范数来近似非凸的 L_0 范数,把组合优化问题转化为凸优化问题。换句话说,通过这样的近似可把一个无法求解的问题转化为一个通过现代优化理论能够求解的问题。这在统计学中也称为 Lasso。

当测量值没有噪声时,得到如下基本追踪表达式:

$$\min \|\mathbf{x}\|_1, \quad \text{s. t. } \mathbf{y} = \Phi \mathbf{x} \quad (3.4.3)$$

当测量值存在少量有界噪声时,得到如下基本追踪去噪方法的表达式:

$$\min \|\mathbf{x}\|_1, \quad \text{s. t. } \|\Phi \mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \quad (3.4.4)$$

2. 无噪声稀疏信号重构

考虑一个更为通用的无噪声稀疏信号重构问题：

$$\mathbf{x}' = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1, \quad \text{s. t. } \mathbf{x} \in B(\mathbf{y}) \quad (3.4.5)$$

其中, $B(\mathbf{y})$ 确保 $\mathbf{x}'$ 与测量值 $\mathbf{y}$ 保持一致。对 $B(\mathbf{y})$ 可有多种选择, 对应式 (3.4.3), $B(\mathbf{y})$ 为 $\{\mathbf{x}: \Phi \mathbf{x} = \mathbf{y}\}$ 。

例如, 假设测量矩阵 Φ 满足 $2K$ 阶约束等距特性且 $\delta_{2K} < \sqrt{2} - 1$ 。已知 $\mathbf{x}, \mathbf{x}' \in \mathbf{R}^N$, 并定义 $\mathbf{h} = \mathbf{x}' - \mathbf{x}$; 设 S 表示矢量 $\mathbf{x}$ 中 K 个幅值最大元素的下标集合, T 表示矢量 $\mathbf{h}_{S^c}$ 中 K 个幅值最大元素的下标集合, $R = S \cup T$ 。如果 $\|\mathbf{x}'\|_1 \leq \|\mathbf{x}\|_1$, 则可以证明

$$\|\mathbf{h}\|_2 \leq C_0 \frac{\sigma_K(\mathbf{x})_1}{\sqrt{K}} + C_1 \frac{|\langle \Phi \mathbf{h}_R, \Phi \mathbf{h} \rangle|}{\|\mathbf{h}_R\|_2} \quad (3.4.6)$$

其中, $\sigma_K(\mathbf{x})_1 = \|\mathbf{x}_{S^c}\|_1 = \|\mathbf{x} - \mathbf{x}_S\|_1$, C_0 和 C_1 都是仅依赖于 δ_{2K} 的常数。

式 (3.4.6) 为式 (3.4.5) 中 L_1 范数最小化方法带来的误差建立了一个界限, 其中的前提条件是测量矩阵 Φ 满足 $2K$ 阶约束等距特性。为了针对具体的 $B(\mathbf{y})$ 构建出特殊的界限条件, 需要考虑 $\mathbf{x}' \in B(\mathbf{y})$ 是如何影响 $|\langle \Phi \mathbf{h}_S, \Phi \mathbf{h} \rangle|$ 的。

以没有测量噪声的情况为例。可以证明, 当 $\mathbf{x} \in \Sigma_K = \{\mathbf{x}: \|\mathbf{x}\|_0 \leq K\}$ 时, 如果 Φ 满足约束等距特性, 则只要 $O(K \ln(N/K))$ 个采样值就可以无失真地重建任何包含 K 个非零元素的目标信号 $\mathbf{x}$, 而不需考虑这 K 个非零元素具体如何分布。

3. 有噪声稀疏信号重构

考虑一个通用的、存在噪声污染情况下的稀疏信号重构问题：

$$\mathbf{x}' = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1, \quad \text{s. t. } \mathbf{x} \in B(\mathbf{y}) \quad (3.4.7)$$

其中, $B(\mathbf{y})$ 确保 $\mathbf{x}'$ 与测量值 $\mathbf{y}$ 保持一致。对 $B(\mathbf{y})$ 可有多种选择, 下面考虑两种情况。

(1) 有界噪声污染信号的重构

如果被污染信号的噪声是有界的, 则假设测量矩阵 Φ 满足 $2K$ 阶约束等距特性且 $\delta_{2K} < \sqrt{2} - 1$ 。假设 $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$, 其中 $\mathbf{e}$ 是测量过程带来的误差。由于噪声是有界的, 即 $\|\mathbf{e}\|_2 \leq \epsilon$ (ϵ 为噪声界), 则当 $B(\mathbf{y}) = \{\mathbf{x}: \|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \epsilon\}$ 时, 可以证明式 (3.4.7) 的解 $\mathbf{x}'$ 满足

$$\|\mathbf{h}\|_2 = \|\mathbf{x}' - \mathbf{x}\|_2 \leq C_0 \frac{\sigma_K(\mathbf{x})_1}{\sqrt{K}} + C_1 \epsilon \quad (3.4.8)$$

其中, C_0 和 C_1 为仅依赖于 δ_{2K} 的常数。

(2) 高斯噪声污染信号的重构

假设 $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$, 噪声 $\mathbf{e} \in \mathbf{R}^M$ 的各分量来自于均值为零、方差为 σ^2 的高斯分布, 即满足标准正态分布 $N(0, \sigma^2 \mathbf{I})$, 则可推出总存在正实数 k , 使得对任何 $\epsilon > 0$ 有下式成立：

$$P(\|\mathbf{e}\|_2 \geq (1 + \epsilon)\sqrt{M}\sigma) \leq \exp(-k\epsilon^2 M) \quad (3.4.9)$$

将式 (3.4.9) 与式 (3.4.8) 结合考虑, 并令 $\epsilon = 1$ 可以推出当 $B(\mathbf{y}) = \{\mathbf{z}: \|\Phi \mathbf{z} - \mathbf{y}\|_2 \leq 2\sigma\sqrt{M}\}$ 时, 式 (3.4.7) 的解 $\mathbf{x}'$ 以至少 $1 - \exp(-kM)$ 的概率满足

$$\|\mathbf{h}\|_2 = \|\mathbf{x}' - \mathbf{x}\|_2 \leq 8 \frac{\sqrt{1 + \delta_{2K}}}{1 - (1 + \sqrt{2})\delta_{2K}} \sqrt{M}\sigma \quad (3.4.10)$$

两种情况下重构都是有界的, 这表明压缩感知确可应用到实际中。

3.4.2 测量矩阵的校准

在讨论具体重构算法前,还有一个问题要先分析一下。如果测量矩阵本身存在噪声,则还需要考虑测量矩阵的校准问题。

回到式(3.4.3)和式(3.4.4),有可能测量矩阵并不是确切已知的,它可能通过模型来描述,但并不是实际中的测量矩阵;或者有些情况下测量矩阵被校准后,随着客观工作条件变化,测量矩阵的物理条件也可能漂移。为解决这类问题,常用以下4种方案。

(1) 忽略这个问题,这有可能明显影响重建精度。

(2) 简单地把由非精确测量矩阵带来的影响当作噪声:假设实际的非精确测量矩阵为 Φ' ,则观测值 $y=\Phi'x+\epsilon$,这样就可把压缩感知的重建问题转换为解下列问题:

$$y = \Phi'x + \epsilon + \eta \quad (3.4.11)$$

其中, η 为由测量矩阵不精确带来的误差。实际中,由于很难获取 η 的统计特性,所以求解式(3.4.11)有一定的难度。

(3) 监督校准:利用已知的训练稀疏信号 $x_1, x_2, \dots, x_l, \dots, x_L$ 和相应的测量值 $y_l = \Phi'x_l + \epsilon_l$ 。把所有的矢量排列起来组成矩阵的形式,有

$$Y = \Phi'X + E \quad (3.4.12)$$

其中,需要基于这些已知的训练信号和测量值来校准测量矩阵:

$$\Phi' := \operatorname{argmin}_{\Phi'} \|Y - \Phi'X\|_F^2 \quad (3.4.13)$$

即要优化 Φ' 以使 $\|Y - \Phi'X\|$ 最小。

这种方案在能够获得已知稀疏信号的情况下可以使用,但在无法获得训练稀疏信号的情况下就无法实用。如果可以假设待校准的测量矩阵来自某些矩阵族也是有用的。例如,有时已知或假设未知的测量矩阵 Φ 在已知的字典中是稀疏的,即 $\Phi \approx \sum_j a_j \Phi_j$,且 $\|a\|_0$ 很小,其中 $a=[a_1, \dots, a_j, \dots]$ 。这样,监督校准问题就可以转化为下面的凸优化问题:

$$\min \|a\|_1 \quad \text{s. t.} \quad \|Y - \sum_j a_j \Phi_j\|_F^2 \leq \epsilon \quad (3.4.14)$$

(4) 非监督校准:针对少数几个未知信号(需要保证具有稀疏的特性),利用它们各自有限的、基于压缩感知模型的采样值,开展盲校准,即无须采用已知的训练信号来校准测量矩阵,或训练的目标稀疏信号是未知的。具体是把未知的训练信号矢量 $x_1, x_2, \dots, x_l$ 表述成矩阵 X ,基于已知精确的测量矩阵 Φ_0 和多组观测矢量 $y_1, y_2, \dots, y_l$ 形成的矩阵 Y ,通过一定的方法可以确定增益矩阵 D 和 X ,即解决下面这个问题:

$$\min_{D, X} \|X\|_1 \quad \text{s. t.} \quad Y = D\Phi_0 X \quad (3.4.15)$$

不过,求解式(3.4.15)有可能导致出现无意义的解。如果对矩阵 D 和 X 没有限制,那么只要把矩阵 D 设为无穷大、让矩阵 X 趋于零,就可满足式(3.4.15),但这样的解并没有实际意义。

这里用一种描述方法重新表述这个问题,以提供一个凸集公式。将 $Y=D\Phi_0 X$ 表述为 $\Delta Y=\Phi_0 X$,其中, $\Delta=D^{-1}=\operatorname{diag}(\delta_i)$, $\operatorname{diag}(\delta_i)$ 是将所有的 δ_i 排列在对角线上形成的对角矩阵。为了避免前述的无意义解 $\Delta=X=0$,引入一个凸集归一化约束条件 $\operatorname{Tr}(\Delta)=\sum_i \delta_i=M$,其中 $\operatorname{Tr}(\cdot)$ 表示矩阵的迹,即 $D=\operatorname{diag}(d_i)$, $\sum_i d_i^{-1}=M$ 。这个约束条件也可以替换为 $\delta_i=1$

等其他形式。由于可以对 Δ 引入任意约束, Δ 和 D 本质上存在一个比例关系。以非监督方式校准测量矩阵的方法可以表示为

$$(\mathbf{X}', \Delta') := \operatorname{argmin} \|\mathbf{X}\|_1 \quad \text{s. t. } \Delta \mathbf{Y} = \Phi_0 \mathbf{X}, \operatorname{Tr}(\Delta) = M \quad (3.4.16)$$

3.4.3 典型重构算法

有效的图像重构算法是压缩感知的关键技术之一。因为在不同的应用中对重构有不同的要求, 所以重构算法和模型的设计多种多样。现在已有了许多不同的重构算法, 它们在运算时间、重构精度和稳定性等方面都有各自的特点。

1. 重构要考虑的因素

设计稀疏信号重构算法需要考虑多方面的因素:

(1) 先验信息: 如何充分发掘图像的先验信息从而构造有效的约束条件是图像重构的关键。目前常用的先验信息主要包括信号稀疏信息以及图像全变分信息。信号稀疏性体现在原始图像信号在某固定变换域或稀疏基(如离散余弦变换基、小波基等)上的投影系数稀疏, 全变分值则考虑了图像相邻像素之间的相关性。

(2) 测量值数量: 要求基于尽可能少的测量, 稳定地重构出包含 K 个非零值的原始稀疏信号。

(3) 抗噪声鲁棒性: 要求无论测量值包含噪声或测量系统本身具有系统噪声, 重构算法都要稳定地重构出原始稀疏信号。

(4) 重构速度: 要求在占用较少计算资源的情况下, 重构算法能高效地实现稀疏信号重构。

(5) 稳定性: 采用 L_1 范数最小化可以确保重建的稳定性, 具体需要考虑相同的重建条件。

通过考虑上述一些或全部因素, 多数重建算法可分成 4 大类, 包括凸优化算法、贪婪算法、组合算法、贝叶斯算法。下面分别对每大类介绍 1~2 个典型算法。

2. 凸优化算法

这类算法也称最优化逼近方法, 其基本思路是用凸函数代替 L_0 范数, 并在一个 $\mathbf{R}^N$ 空间中的凸集里去优化关于未知变量 $\mathbf{x}$ 的凸目标函数 $J(\mathbf{x})$ 。这类算法重构精度高, 需要的测量数据比较少, 约为 $O(K \log(N/K))$; 但计算速度慢, 计算复杂性约为 $O(N^3)$ 。

假设 $J(\mathbf{x})$ 是一个能促进稀疏性的(即当目标信号 $\mathbf{x}$ 很稀疏时, $J(\mathbf{x})$ 的值很小)且凸的代价函数。在测量值没有噪声时, 基于测量值 $\mathbf{y} = \Phi \mathbf{x}$ 要重建信号 $\mathbf{x}$ 可以解下面的方程:

$$\min\{J(\mathbf{x})\}, \quad \text{s. t. } \mathbf{y} = \Phi \mathbf{x} \quad (3.4.17)$$

而在测量值有噪声时, 解下面的方程:

$$\min\{J(\mathbf{x})\}, \quad \text{s. t. } H(\Phi \mathbf{x}, \mathbf{y}) \leq \epsilon \quad (3.4.18)$$

其中, H 为用来对 $\Phi \mathbf{x}$ 和 $\mathbf{y}$ 之间的矢量距离进行惩罚的代价函数。

式(3.4.18)也可以改写成没有约束条件的形式:

$$\min\{J(\mathbf{x}) + \lambda H(\Phi \mathbf{x}, \mathbf{y})\} \quad (3.4.19)$$

其中, λ 为一个惩罚因子, 它可以通过试错法或交叉验证法来选取。常用的 J 和 H 为: $J(\mathbf{x}) = \|\mathbf{x}\|_1$, 即为稀疏信号 $\mathbf{x}$ 的 L_1 范数; $H(\Phi \mathbf{x}, \mathbf{y}) = 0.5 \times \|\Phi \mathbf{x} - \mathbf{y}\|_2^2$, 即为实际测量值 $\mathbf{y}$ 和理论测量值 $\Phi \mathbf{x}$ 之间误差的 L_2 范数。在统计学中, 在 $\|\mathbf{x}\|_1 \leq \delta$ 的条件下最小化 $H(\Phi \mathbf{x}, \mathbf{y})$ 称为 Lasso 问题。

(1) 总变分重构法

从广义的角度, $J(\cdot)$ 作为一个正则项也可以是其他复杂函数。例如, 如果目标信号是阶梯函数, 同时又在一个已知变换域 Φ 中体现出稀疏性, 则可以有一个如下的混合正则项:

$$J(\mathbf{x}) = \text{TV}(\mathbf{x}) + \lambda \|\Phi\mathbf{x}\|_1 \quad (3.4.20)$$

其中, $\text{TV}(\mathbf{x})$ 为目标信号的总变分(所以得名):

$$\text{TV}(\mathbf{x}) = \sum_{i=0}^N |x_{i+1} - x_i| \quad (3.4.21)$$

可见, 针对阶梯函数类型(或有丰富边缘)的信号 $\mathbf{x}$, 总变分 $\text{TV}(\mathbf{x})$ 将凸显出稀疏性。

(2) 收缩循环迭代法

如果考虑式(3.4.19)中 H 是二次型(凸函数且可导), 常可以使用收缩循环迭代法, 也称软阈值法来解。该方法原是一种经典的基于小波变换域的图像去噪声方法, 其中的收缩运算符定义为

$$\text{shrink}(t, a) = \begin{cases} t - a & t > a \\ 0 & -a \leq t \leq a \\ t + a & t < -a \end{cases} \quad (3.4.22)$$

用收缩循环迭代法解决式(3.4.19)的问题也很有效, 基本方法可写成定点迭代的方式: 对于 $i=1, \dots, N$, 可将目标信号 $\mathbf{x}$ 的第 i 个分量在第 $k+1$ 次迭代步骤写成

$$\mathbf{x}_i^{k+1} = \text{shrink}\left[\mathbf{x}_i^k - \tau \frac{\partial H(\mathbf{x}_i^k)}{\partial \mathbf{x}_i^k}, \mu\tau\right] \quad (3.4.23)$$

其中, $\tau > 0$ 是梯度下降法的步长, 它可随着 k 而变化; μ 可以根据噪声大小和经验来选取, μ 越大说明允许 $\mathbf{x}_i^{k+1}$ 和 $\mathbf{x}_i^k$ 之间的距离越大。

针对有噪声污染测量值的重构问题, 一种选取凸优化代价函数 $H(\Phi\mathbf{x}, \mathbf{y})$ 的常规方法是采用残差的 L_2 范数:

$$H(\Phi\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \quad \frac{\partial H}{\partial \mathbf{x}} = \Phi^T(\Phi\mathbf{x} - \mathbf{y}) \quad (3.4.24)$$

针对这个特殊的惩罚函数, 式(3.4.23)可以简化为

$$\mathbf{x}_i^{k+1} = \text{shrink}[\mathbf{x}_i^k - \tau \Phi^T(\Phi\mathbf{x}_i^k - \mathbf{y}), \mu\tau] \quad (3.4.25)$$

3. 贪婪算法

从本质上来说, 稀疏信号重构是基于线性测量值 $\mathbf{y}$ 来重构出最具稀疏性的目标信号 $\mathbf{x}$, 即重建出具有最少的非零个数的目标信号 $\mathbf{x}$ 。换句话说, 是求解

$$\min \left\{ |\mathbf{I}| : \mathbf{y} = \sum_{i \in \mathbf{I}} \Phi_{:,i} x_i \right\} \quad (3.4.26)$$

其中, $\mathbf{I} \subseteq \{1, \dots, N\}$, 表示一个索引集(对应支撑集); $\Phi_{:,i}$ 表示矩阵 Φ 的第 i 列。

式(3.4.26)可借助经典的稀疏逼近方法通过逐步选择矩阵 Φ 的列来逼近 $\mathbf{y}$, 进而逐步确定索引集 $\mathbf{I}$ 。

这类算法的复杂度大多是由寻找到正确索引集所需要的迭代次数所决定的, 计算速度一般比较快, 但是需要的测量数据多且重构的精度比较低。

(1) 匹配追踪算法

匹配追踪(MP)算法是一种基于迭代的贪婪算法。匹配追踪将一个信号看成由基/字典中的元素通过线性组合而构成的。在压缩感知的稀疏重建中, 字典就是采样矩阵 $\Phi \in$

$\mathbf{R}^{M \times N}$, 它的每列表示一种原型信号的元素(也称原子信号)。重构可描述为: 给定一个信号 $\mathbf{y}$, 寻求通过这些基/字典中的元素的稀疏性组合来描述信号 $\mathbf{y}$ 。

匹配追踪算法主要是围绕原始观测信号 $\mathbf{y}$ 与线性组合的残差 $\mathbf{r} \in \mathbf{R}^M$ 展开, 残差主要描述没有被解释的测量值。在每次迭代中, 从字典里选取一个跟残余分量差相关性最大的列:

$$\lambda_k = \operatorname{argmin}_k \left\{ \frac{\langle \mathbf{r}_{k-1}, \boldsymbol{\phi}_{i_k} \rangle}{\|\boldsymbol{\phi}_{i_k}\|^2} \right\} \quad (3.4.27)$$

其中, $\boldsymbol{\phi}_{i_k}$ 为矩阵 $\boldsymbol{\Phi}$ 的第 i_k 列。一旦这个列被选中, 就获得了一个更为逼近原始信号的结果。这是因为一个新的系数 λ_k 被增加到对原始信号逼近的索引集中。

然后, 进行如下更新:

$$\begin{cases} \mathbf{r}_k = \mathbf{r}_{k-1} - \frac{\langle \mathbf{r}_{k-1}, \boldsymbol{\phi}_{i_k} \rangle}{\|\boldsymbol{\phi}_{i_k}\|^2} \boldsymbol{\phi}_{i_k} \\ \mathbf{x}^k = \mathbf{x}^{k-1} + \langle \mathbf{r}_{k-1}, \boldsymbol{\phi}_{i_k} \rangle \end{cases} \quad (3.4.28)$$

经过多次迭代, 残差将变得越来越小, 直到残差的范数小于某个预定的阈值, 算法停止。

上述算法有两个缺点: 一方面, 它不能保证重建误差足够小; 另一方面, 它常需要大量的循环次数才能逼近原始信号。这是因为如果将残差对已选择的元素进行垂直投影是非正交的, 则每次迭代的结果只是次优而不是最优, 所以收敛需要很多次迭代。

(2) 正交匹配追踪算法

在正交匹配追踪(OMP)算法中, 每次循环时都把残差 $\mathbf{r}$ 投影到与所有已选定的列线性展开的正交子空间中, 而不是将残差减去一个与其最大相关的字典中的矢量。由于残差 $\mathbf{r}$ 总是和已选取的列正交, 所以相同的列在 OMP 中不会被选中两次, 因而它的最大循环次数会明显减少。

假设 $\boldsymbol{\Phi}_k$ 表示矩阵 $\boldsymbol{\Phi}$ 在第 k 个步骤中选择出的子矩阵(第 k 个列的选取过程与 MP 相同), 则

$$\mathbf{x}_k = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}_k \mathbf{x}\| \quad \mathbf{r}_k = \mathbf{y} - \boldsymbol{\Phi}_k \mathbf{x}_k \quad (3.4.29)$$

可以证明, 如果原始的待测量信号是稀疏的, 同时测量矩阵中每个元素都是从一个服从亚高斯分布的变量中随机选取的, 则基于线性混叠的有限个测量值, 可通过 OMP 以极大的概率重建出原始的稀疏信号。这个算法最多需要 K 次循环即可收敛, 其中 K 是原始信号中的非零个数。该算法可保证每次迭代的最优性, 从而减少了迭代次数。但它每次迭代中仅选取一个元素来更新已选的元素集合, 迭代的次数与稀疏度 K 或采样个数 M 密切相关。另外, 每次迭代还需要额外的正交化运算, 随着 K 或 M 的增加, 运算时间也会大幅增加。OMP 的运算复杂度为 $O(MNK)$ 。

(3) 分段匹配追踪算法

对 OMP 的一种改进是分段正交匹配追踪(StOMP)。它与 OMP 在每次循环里仅从字典中选取一个元素不同, 它每次选取一个元素集合, 该集合的特点是残余分量与字典列的相关性均大于一定的阈值, 而后残余分量将更新字典的列集合。StOMP 的运算复杂度为 $O(MN \ln N)$, 远比 OMP 小。

(4) 子空间追踪算法

子空间追踪(SP)算法也是对 OMP 的一种改进。它能在无噪声和有噪声时分别地进行精确的信号重建和逼近的信号重建。对任意测量矩阵, 只要它满足约束等距特性, 那么 SP

就能够从无噪声的测量矢量中精确地重建原始信号。

在压缩感知信号重建算法中,最重要的就是确定测量矢量 $\mathbf{y}$ 位于哪个子空间,且这些子空间是由测量矩阵中哪 K 个列矢量生成的。一旦确定了正确的子空间位置,通过应用子空间的伪逆运算就可以计算出信号的非零系数。SP 的重要特点就是要寻找生成正确子空间的 K 个列矢量。该算法包含测量矩阵中的 K 个列矢量的列表,对子空间的最初估计是测量矩阵中与实际信号 $\mathbf{y}$ 的 K 个最大相关的列。为了修正对子空间的最初估计,SP 会检测包含 K 个列矢量的子集,检测这个子空间是否可以很好地重建信号。假如重建信号与实际信号之间的残差没有达到算法的要求则需要更新这个列表。该算法会通过保留可靠的、丢弃不可靠的候选值,同时也加入相同数量的新候选值来更新子空间。更新的原则是新的子空间重建信号残差要小于更新前的子空间重建信号残差。该算法在一定条件下能够确保重建,且可以确保下一个迭代循环能够找到更好的子空间。

4. 组合算法

组合算法的基本思想是先对信号进行高度结构化采样(线性投影),再借助组/群测试以快速获得信号的支撑集,实现精确重构。将重构问题表述成:已知长度为 N 的矢量 $\mathbf{x}$ 中包含 K 个非零元素, $\mathbf{x}_i \neq 0$,但它们的位置分布未知。要求设计一种测试方法,通过最少的测试次数确定非零元素所在的位置。最简单实用的方法是把测试表示成一个二进制矩阵 Φ ,其元素 $\phi_{i,j}=1$ 表示在第 j 次测试中,第 i 个元素被选用。如果输出信号与输入信号是一种线性关系,则重构目标稀疏矢量 $\mathbf{x}$ 的问题就可转化为标准的压缩感知稀疏重构问题。组合算法需要的测量次数较少,运算速度高,但是一般测量矩阵复杂,且不易确定对测量矩阵的约束条件,这为构建新的测量矩阵带来了困难。

下面介绍两个简单的组合算法,均要求测量值没有噪声干扰。

(1) 计数-最小略图法

计数-最小略图法仅考虑非负信号。设 H 表示所有离散值函数 $h: \{1, \dots, N\} \rightarrow \{1, \dots, m\}$ 的集合, H 是一个有限集合,其大小为 m^N 。 H 中每一个函数 h 可以通过一个大小为 $m \times N$ 的二进制矩阵 $\Phi(h)$ 来表示,其中每一列都是一个二进制矢量,每一列仅在位置 $j = h(i)$ 包含一个 1。为了构建完整的采样矩阵 Φ ,可以根据均匀分布从 H 中独立选取 d 个函数 $h_1, \dots, h_d$, 把它们的二进制矩阵垂直堆叠成一个尺寸为 $M \times N$ 的矩阵,该矩阵即为 Φ ,且其每列均有 d 个 1,其中 $M = md$ 。

针对已知任意信号 $\mathbf{x}$,可以获得线性测量值 $\mathbf{y} = \Phi \mathbf{x}$ 。根据下面的两个特性,很容易得到对测量值的一个直观认识。首先,作为测量值矢量的 $\mathbf{y}$ 很容易借助母二进制函数 $h_1, \dots, h_d$ 得到分组的形式;其次,针对测量矢量 $\mathbf{y}$ 的第 i 个系数与母函数 h 有关,即

$$y_i = \sum_{j: h(j)=i} x_j \quad (3.4.30)$$

换句话说,对一个固定的信号系数 j ,每个测量值 y_i 都有函数 h 把 x_j 以汇总的形式映射到相同的 i 上。信号重构的目标就是从这些汇总后的观察值 y_i 中,恢复出原始的信号值 x_j 。

当原始信号是正数时,计数-最小略图法非常有用。已知测量值 $\mathbf{y}$ 和采样矩阵 Φ ,重建目标信号的第 j 个系数 x'_j 可以由下式完成:

$$x'_j = \min_i y_i: h_i(j) = i \quad (3.4.31)$$

直观地说,这就意味着重建目标信号的第 j 个系数 x'_j 可以通过从所有可能有 x_j 介入而形成的测量值中,找出其幅值最小的一个观测值(作为 x'_j)来得到。该方法明显的特点是简单高效。

(2) 计数-中值略图法

计数-中值略图法在目标信号有可能为负数时也适用。这里不是取幅值最小的一个观测值作为 x'_j ,而是取中值。因为对一个普通的信号,其他的 x 对 y 的影响可能是正的也可能是负的,而中值最有可能是原始信号的值。已知测量值 y 和采样矩阵 Φ ,重建目标信号的第 j 个系数 x'_j 可以由下式完成:

$$x'_j = \text{median}_i y_i: h_i(j) = i \quad (3.4.32)$$

5. 贝叶斯算法

前述各种方法都认为信号是确定的或属于某个已知集合,即在已知信号模型的前提下讨论重构。贝叶斯压缩感知的基本思想是在未知信号模型的基础上考虑非确定性因素,即考虑一个概率分布已知的稀疏信号,从随机测量中重构符合此概率分布的信号。该类算法考虑了信号之间的时间相关性,所以对具有较强时间相关性的信号可提供比其他重构算法更高的重构精度。不过,由于在贝叶斯信号建模的框架下没有确切的“无失真”或“重建误差”的概念,所以下面介绍的各种算法与前述算法不同,并不能基于一定数量的测量值无失真地重建原始目标信号。

(1) 相关向量机

相关向量机(RVM)是一种贝叶斯学习方法。它能够产生稀疏的分类结果,其机理是从许多待选的基函数中选择少数几个,用它们的线性组合作为分类函数来实现分类。从压缩感知的角度出发,可将其视为一种确定稀疏信号中分量的方法,这种信号给一些基函数提供了不同的权重,而这些基函数就是测量矩阵 Φ 中的列向量。

相关向量机使用了几个层次的先验概率模型。首先,假设 x 的每个分量都独立并服从均值为零而方差待定的高斯分布,即

$$p(x_i | \alpha_i) = N(x_i | 0, \alpha_i^{-1}) \quad (3.4.33)$$

其中, $N(x | m, \sigma^2)$ 为随机变量 x 服从均值为 m 、方差为 σ^2 的高斯分布。

进一步,有

$$p(\mathbf{x} | \boldsymbol{\alpha}) = \prod_{i=1}^N N(x_i | 0, \alpha_i^{-1}) \quad (3.4.34)$$

其中, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]$ 。

然后,假设高斯分布中方差的逆(也就是 $\boldsymbol{\alpha}$)也独立且服从参数为 a, b 的伽马分布

$$p(\boldsymbol{\alpha} | a, b) = \prod_{i=1}^N \text{Gamma}(\alpha_i | a, b) \quad (3.4.35)$$

其中, $\text{Gamma}(\alpha_i | a, b) = \alpha_i^{a-1} e^{-a\alpha_i} \Gamma(a)^{-1} b^a$, 且 $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ 。

可见,使用方差的逆可以控制赋予某个分量的权重。通过使用贝叶斯推理,可以得到在给定参数 a 和 b 的条件下, x 的边缘分布是一个未标准化学学生氏分布,即

$$p(x_i | a, b) = \int p(x_i | \alpha_i) p(\alpha_i | a, b) d\alpha_i = \frac{b^a \Gamma(a + \frac{1}{2})}{\sqrt{2\pi} \Gamma(a)} \left[b + \frac{x_i^2}{2} \right]^{-(a + \frac{1}{2})} \quad (3.4.36)$$

未标准化学生氏分布是由通过拉伸和平移服从标准化学生氏分布的随机变量得来的。标准化学生氏分布为

$$p(t | v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left[1 + \frac{t^2}{v}\right]^{-\left(\frac{v+1}{2}\right)} \quad (3.4.37)$$

随机变量 $r = \mu + st$ 的分布就是未标准化学生氏分布：

$$p(t | v, \mu, s) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{s\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left[1 + \frac{(r-\mu)^2}{vs^2}\right]^{-\left(\frac{v+1}{2}\right)} \quad (3.4.38)$$

可以看到, $\mathbf{x}$ 的边缘分布是参数 $\mu=0, v=2a$ 和 $s=\sqrt{(b/a)}$ 的未标准化学生氏分布。学生氏分布可以促进稀疏性的产生。如果假定误差服从均值为零方差为 σ^2 的高斯分布, 即

$$\mathbf{y} = \Phi \mathbf{x} + \varepsilon \quad \varepsilon \sim N(0, \sigma^2) \quad (3.4.39)$$

则给定测量值 $\mathbf{y}$, 就可以通过贝叶斯推理结合使用各种迭代算法来获得 $\mathbf{x}$ 的后验概率分布。

(2) 贝叶斯压缩感知

可从 RVM 模型出发来考虑贝叶斯压缩感知(BCS)。对于给定观测值 $\mathbf{y}$, 可以通过将 $\mathbf{y}$ 给定 $\mathbf{x}$ 的概率分布对 $\mathbf{x}$ 积分直接求出其边缘对数似然率, 进而使用 EM 算法(见 12.5.2 小节)求解各个参数。由式(3.4.39)可以得到

$$p(\mathbf{y} | \mathbf{x}, \boldsymbol{\alpha}, \sigma^2) = N(\mathbf{y} - \Phi \mathbf{x}, \sigma^2 \mathbf{I}) \quad (3.4.40)$$

其中, $\mathbf{I}$ 为大小合适(这里是 $M \times N$)的单位矩阵。借助贝叶斯推理, 可以得到

$$\begin{aligned} p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2) &= \int p(\mathbf{y} | \mathbf{x}, \sigma^2) p(\mathbf{x} | \boldsymbol{\alpha}) d\mathbf{x} \\ &= (2\pi)^{-N/2} |\sigma^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{y}^T (\sigma^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T)^{-1} \mathbf{y}\right] \end{aligned} \quad (3.4.41)$$

其中, $\mathbf{A}$ 是一个对角矩阵, 其对角线上的元素由 $\boldsymbol{\alpha}$ 组成。

在这个过程中, 需要对一个 $N \times N$ 的矩阵求逆, 所以算法复杂度是 $O(N^2)$ 。采用快速边缘似然率最大化的方法可以将复杂度降低到 $O(NM^2)$ 。它是一种渐进式的模型构造法, 将基函数顺序地加入或删除, 所以可极大地利用信号的稀疏性。

贝叶斯压缩感知通过逐步迭代的方法逼近待重构稀疏信号的支撑集, 它的一个优点是它可以对所估计的信号的每一个分量提供一个置信区间。这样就可帮助了解和判定: 是否该估计是准确的。例如, 太大的置信区间表明这个估计可能不是十分可靠。理想情况下, 希望得到一个估计值, 同时其置信区间很小。

3.5 稀疏编码与字典学习

稀疏编码与字典学习的结合曾与压缩感知并行发展, 密切相关又互相影响。稀疏编码源自观察: 某类信号多是由少数几个基本原子的加权组合而构成。这正对应信号的稀疏表达。字典学习的目标就是要估计出给定目标信号的原子信号及其权重, 从而实现对目标信号的稀疏表达。

可以采用数学模型来描述字典学习。为简便, 这里仅考虑实数信号。设 $\mathbf{y}_i \in \mathbf{R}^L$ 是一个

维度为 L 的矢量,用来表示一共 N 个信号中的第 i 个; $\mathbf{A}=[a_1, \dots, a_M] \in \mathbf{R}^{L \times M}$ 表示原子信号矩阵,其中第 i 列是维度为 L 的原子信号; $\mathbf{w}_i \in \mathbf{R}^M$ 是一个维度为 M 的矢量,用来表示构成矢量 $\mathbf{y}_i$ 的原子信号的权重矢量。根据前面的假设,有

$$\mathbf{y}_i = \mathbf{A}\mathbf{w}_i + \boldsymbol{\epsilon}_i \quad |\mathbf{w}_i|_0 \leq K \quad i = 1, \dots, N \quad (3.5.1)$$

其中, $\boldsymbol{\epsilon}_i \in \mathbf{R}^L$ 为来自某个概率分布的未知误差; $|\mathbf{x}|_0$ 为矢量 $\mathbf{x}$ 的 L_0 范数; K 为一个大于零的正整数。

式(3.5.1)表达了多个重要的信息:

(1) 信号是原子矢量的线性组合,所以这里使用了线性模型。线性模型的优点是简单、容易理解和操作。其缺点是不易描述具有非线性的数据。不过,这里并没有对信号所在的空间加以限制,所以如果要引入非线性成分,可将信号转换到某个特征空间后再使用线性模型。

(2) 非确定性是由误差项 $\boldsymbol{\epsilon}_i$ 来体现的。这里并未假设误差来自何种概率分布,虽然常用的是高斯分布。如果希望原子信号是非相关的,且假设误差分布是均值为零、方差为单位矩阵的多维高斯分布,就可得到 PCA 模型。

(3) 原子信号的权重矢量 $\mathbf{w}_i$ 是 K 稀疏的,这是由 $|\mathbf{w}_i|_0 \leq K$ 表达的。这表明感兴趣的信号是由原子信号构成的。需要指出,稀疏性假设并非在字典学习中不可或缺。例如,盲源分离的典型方法,如独立成分分析和因素分析(FA)都是没有稀疏性假设的线性模型,也可被归到字典学习的范畴中。

可将式(3.5.1)表达为矩阵形式

$$\mathbf{Y} = \mathbf{A}\mathbf{W} + \mathbf{E} \quad |\mathbf{w}_i|_0 \leq K \quad i = 1, \dots, N \quad (3.5.2)$$

其中, $\mathbf{Y}=[\mathbf{y}_1, \dots, \mathbf{y}_N]$, $\mathbf{W}=[\mathbf{w}_1, \dots, \mathbf{w}_N]$, $\mathbf{E}=[\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N]$ 。字典学习的目的是从数据矩阵 $\mathbf{Y}$ 中学习,即求解 $\mathbf{A}$ 和 $\mathbf{W}$ 。

压缩感知的出发点是已知待测信号是稀疏的,需设计测量矩阵以得到测量值,再从测量值中推出待测信号。所以,在压缩感知的框架中已知测量矩阵和测量值,未知参数只有一个,即待测信号,而且待测信号只有一个样本。但在字典学习中,仅仅知道很多观测信号,而未知参数却有两个,即原子信号矩阵 $\mathbf{A}$ 及权重矩阵 $\mathbf{W}$,并且需要考虑多个样本。可见,字典学习更为困难,求解也更复杂。

3.5.1 字典学习与矩阵分解

这里先忽略稀疏性假设,集中讨论字典学习。如果忽略误差,式(3.5.2)具有如下简单形式:

$$\mathbf{Y} \approx \mathbf{A}\mathbf{W} \quad (3.5.3)$$

其中,已知量是 $\mathbf{Y}$,即众多信号的观察值,需要计算的是 $\mathbf{A}$ 和 $\mathbf{W}$ 。所以,式(3.5.3)是一个矩阵乘式分解的问题。例如可以利用 SVD 将 $\mathbf{Y}$ 写成

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (3.5.4)$$

其中, $\mathbf{U} \in \mathbf{R}^{L \times L}$ 为一个包含正交单位列向量的矩阵,即 $(i, j=1, \dots, L)$

$$\mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (3.5.5)$$

其中, $\mathbf{u}_i$ 为矩阵 $\mathbf{U}$ 的第 i 列。

$\mathbf{D} \in \mathbf{R}^{L \times N}$ 为一个非主对角线上的元素为零、主对角线的元素全部或部分大于零的矩阵，称为矩阵 $\mathbf{Y}$ 的奇异值矩阵。因为在绝大多数情况下维数 L 不等于 N ，所以矩阵 $\mathbf{D}$ 不一定为方阵。在这种情况下，取 $\mathbf{D}$ 左上角开始的最大子方阵的对角线为其主对角线。因此，

$$\begin{cases} \text{如果 } i = j & \text{则 } d_{ij} \geq 0 \\ \text{如果 } i \neq j & \text{则 } d_{ij} = 0 \end{cases} \quad (3.5.6)$$

其中， d_{ij} 为 $\mathbf{D}$ 中第 i 行第 j 列元素， d_{ii} 就是矩阵 $\mathbf{Y}$ 的第 i 个奇异值，并且如果 $i > j$ ，有 $d_{ii} \geq d_{jj}$ ，也就是主对角线上的奇异值按降序排列。

$\mathbf{Y} \in \mathbf{R}^{N \times N}$ 同 $\mathbf{U}$ 一样是一个包含正交单位列向量的矩阵。

将矩阵 $\mathbf{Y}$ 的秩记为 $R(\mathbf{Y})$ ，它小于或等于 $\min\{L, N\}$ ，即 $R(\mathbf{Y}) \leq \min\{L, N\}$ 。在 SVD 分解中，如果已知 $\mathbf{U}$ 和 $\mathbf{V}$ 都是满秩矩阵，则由 $R(\mathbf{Y}) = R(\mathbf{U}\mathbf{D}\mathbf{V}^T)$ 可知，奇异值矩阵 $\mathbf{D}$ 中的非零奇异值的个数对应矩阵 $\mathbf{Y}$ 的秩。设 $R(\mathbf{Y}) = T$ ，那么可以截掉 SVD 分解中各矩阵中冗余的列，得到稍微简洁的 SVD 分解矩阵，即 $\mathbf{U} \in \mathbf{R}^{L \times T}$ ， $\mathbf{D} \in \mathbf{R}^{T \times T}$ ，并且 $\mathbf{V} \in \mathbf{R}^{N \times T}$ ，通常称为瘦 SVD。

比较式 (3.5.3) 和式 (3.5.4) 可见，如果令 $\mathbf{A} = \mathbf{U}$ ， $\mathbf{W} = \mathbf{V}\mathbf{D}^T$ ，则可得到式 (3.5.3) 的字典学习问题的一个解。但实际上式 (3.5.3) 中是一个约等号，说明有误差存在。而式 (3.5.4) 的 SVD 解没有考虑误差，所以并不符合实际情况。解决的方法是保留若干个最重要的奇异值，将不重要的奇异值设为零。这就得到截断 SVD (tSVD)。

矩阵的秩决定了非零奇异值的个数。重写式 (3.5.4) 如下：

$$\mathbf{Y} = \sum_{i=1}^N d_i \mathbf{u}_i \mathbf{v}_i^T \quad (3.5.7)$$

其中， d_i 为矩阵 $\mathbf{Y}$ 的第 i 个奇异值，且依次从大到小排列 (当 $i > j$ 时， $d_i \geq d_j$)。

tSVD 保留最大的几个奇异值 (设保留了 T 个)，舍弃其余的奇异值来逼近矩阵 $\mathbf{Y}$ ：

$$\mathbf{Y}' = \sum_{i=1}^T d_i \mathbf{u}_i \mathbf{v}_i^T \quad (3.5.8)$$

所以， $\mathbf{Y}'$ 只是一个估计值，它与原来矩阵之间的残差 $\mathbf{R} = \mathbf{Y} - \mathbf{Y}'$ 为

$$\mathbf{R} = \sum_{i=T+1}^N d_i \mathbf{u}_i \mathbf{v}_i^T \quad (3.5.9)$$

所以在采用 tSVD 方法的字典学习方案中可将误差显式地表示出来， $\mathbf{A} = [\mathbf{u}_1, \dots, \mathbf{u}_N]$ ， $\mathbf{W} = [d_1 \mathbf{v}_1, \dots, d_T \mathbf{v}_T]$ 。为确定式 (3.5.9) 的残差在原始信号矩阵 $\mathbf{Y}$ 中所占的比例，可借助矩阵的范数。 $\mathbf{F}$ 范数是矩阵所有元素的平方和，所以

$$\|\mathbf{R}\|_{\mathbf{F}}^2 = \sum_{i=T+1}^N d_i^2 \quad (3.5.10)$$

$$\|\mathbf{Y}\|_{\mathbf{F}}^2 = \sum_{i=1}^N d_i^2 \quad (3.5.11)$$

这样，残差在原始信号矩阵 $\mathbf{Y}$ 中所占的比例为

$$\frac{\|\mathbf{R}\|_{\mathbf{F}}^2}{\|\mathbf{Y}\|_{\mathbf{F}}^2} = \frac{\sum_{i=T+1}^N d_i^2}{\sum_{i=1}^N d_i^2} \quad (3.5.12)$$

可见，在 $\mathbf{F}$ 范数衡量下的残差与原始信号之间的比例差别完全取决于原始信号奇异值平方的分布以及使用多少个 (T) 最大奇异值去重建信号。

为确定 T , 可根据 SVD 与 PCA 之间的关系, 即 SVD 的 U 矩阵对应 PCA 中的主分量, DV^T 就是数据在主分量上的投影值。这样就可看出, 式(3.5.8)的 tSVD 逼近实际上是要选取若干个主分量来重构信号。所以, 可以利用 PCA 中确定主分量个数的方法来确定 tSVD 中的 T (即保留的最大奇异值的个数)。

3.5.2 非负矩阵分解

tSVD 的优化形式可用下式来表示 (第一行是要优化的目标, 下面两行是限定条件):

$$\begin{cases} \min_{A \in \mathbf{R}^{L \times M}, W \in \mathbf{R}^{M \times N}} \|E\|_F^2 \\ \text{s. t. } Y = AW + E \\ A^T A = I_M \end{cases} \quad (3.5.13)$$

其中, I_M 为 $M \times M$ 的单位矩阵。

式(3.5.13)取数据矩阵 Y 的 SVD 分解在 U 矩阵中与 T 个最大奇异值相关的列作为原子信号矩阵 A , 并要求原子信号矩阵 A 正交。

式(3.5.13)有两个问题。首先, tSVD 是它的一个解, 但由于没有限制 W 的形式, 所以式(3.5.13)可能有无穷多个解。事实上, 任何解的西变换都是合法解, 即如果假设 A^* 和 W^* 是一组解, H 是一个 $M \times M$ 的西矩阵 ($H^T H = HH^T = I_M$), 则 $A^* H$ 和 $W^* H$ 也是式(3.5.13)的解。其次, 这里对字典信号必须正交的要求过于苛刻。

从式(3.5.13)可见, 要去掉原子信号必须正交的限制, 只要简单去掉第 3 行就可。但这样可能导致解的空间太大以及平凡解的产生。所以需要对 A 或 W 或两者同时进行适当地限制。这样用来限定解空间的规则应满足两个条件: 一个是限定规则本身应具有物理意义; 另一个是限定后优化问题应可解, 或者说其近似解可在有限时间内计算出来。

字典学习的本质是矩阵分解。非负矩阵分解(NMF)是一种非线性的维数约减手段, 可以满足上面两个条件。实际信号有物理意义, 所以是非负的, 生成这些信号的原子信号也应非负, 权重矩阵要帮助构成信号所以也应非负。NMF 的限定条件就是原子信号和权重矩阵必须都是非负的。NMF 的优化形式类似于式(3.5.13), 可写为

$$\begin{cases} \min_{A \in \mathbf{R}^{L \times M}, W \in \mathbf{R}^{M \times N}} \|E\|_F^2 \\ \text{s. t. } Y = AW + E \\ A \geq 0 \\ W \geq 0 \end{cases} \quad (3.5.14)$$

考虑到优化目标的多样性, 还可写出更具有普适性的形式

$$\begin{cases} \min_{A \in \mathbf{R}^{L \times M}, W \in \mathbf{R}^{M \times N}} d(Y, AW) \\ \text{s. t. } Y = AW + E \\ A \geq 0 \\ W \geq 0 \end{cases} \quad (3.5.15)$$

其中, $d(P, Q)$ 为表示矩阵 P 和 Q 之间非相似度的函数。

如果设

$$d(\mathbf{P}, \mathbf{Q}) = \|\mathbf{P} - \mathbf{Q}\|_{\text{F}}^2 \quad (3.5.16)$$

则式(3.5.15)退化为式(3.5.14)的形式。

NMF 的解空间一般还偏大,对它的改进常常是增加一些限制条件以进一步收缩解空间。例如,可以加入权重矩阵的稀疏性条件以实现稀疏非负矩阵分解:

$$\begin{cases} \min_{\mathbf{A} \in \mathbf{R}^{L \times M}, \mathbf{W} \in \mathbf{R}^{M \times N}} d(\mathbf{Y}, \mathbf{AW}) \\ \text{s. t. } \mathbf{Y} = \mathbf{AW} + \mathbf{E} \\ \mathbf{A} \geq 0 \\ \mathbf{W} \geq 0 \\ \|\mathbf{W}\|_0 \leq K \end{cases} \quad (3.5.17)$$

另外,对流形上的非负矩阵分解可以有

$$\begin{cases} \min_{\mathbf{A} \in \mathbf{R}^{D \times M}, \mathbf{W} \in \mathbf{R}^{M \times N}} d(\mathbf{Y}, \mathbf{AW}) \\ \text{s. t. } \mathbf{Y} = \mathbf{AW} + \mathbf{E} \\ \mathbf{A} \geq 0 \\ \mathbf{W} \geq 0 \\ \mathbf{a}_i \in S \quad i = 1, \dots, S \end{cases} \quad (3.5.18)$$

其中, S 为具有特定结构的流形。

3.5.3 端元提取

端元提取也是一类字典学习方法,其两个限制条件为:一个是权重矩阵非负;另一个是权重矩阵每一列的和必须为 1,即

$$\mathbf{y}_i = \mathbf{A}\mathbf{w}_i + \mathbf{e} \quad \mathbf{w}_i \geq 0 \quad \mathbf{w}_i^T \mathbf{L} = 1 \quad (3.5.19)$$

其中, $\mathbf{L}$ 为一个长度合适的全为 1 的列向量。

将式(3.5.19)写成矩阵形式为

$$\mathbf{Y} = \mathbf{AW} + \mathbf{E} \quad \mathbf{W} \geq 0 \quad \mathbf{W}^T \mathbf{L} = \mathbf{L} \quad (3.5.20)$$

一种典型的端元提取优化方法是迭代约束端元化(ICE):

$$\begin{cases} \min_{\mathbf{A} \in \mathbf{R}^{L \times M}, \mathbf{W} \in \mathbf{R}^{M \times N}} \|\mathbf{E}\|_{\text{F}}^2 + \lambda \sum_{i \neq j} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 \\ \text{s. t. } \mathbf{Y} = \mathbf{AW} + \mathbf{E} \\ \mathbf{W} \geq 0 \\ \mathbf{W}^T \mathbf{L} = \mathbf{L} \end{cases} \quad (3.5.21)$$

其中, $\mathbf{A}$ 为端元矩阵,其第 i 列,即 $\mathbf{a}_i$ 是第 i 个端元,求和项表示两两不同端元之间的距离和; λ 为正则参数,用以权衡数据重建误差与距离和。

ICE 的模型和优化问题与非负矩阵分解没有本质区别,但优化过程很不一样。ICE 采用的是类似于坐标交替下降迭代的优化方法。在某一个迭代步骤 t ,固定一个未知量,如 $\mathbf{W}^t$,去求解另一个未知量,如 $\mathbf{A}^{t+1}$;得到 $\mathbf{A}^{t+1}$ 后再固定 $\mathbf{A}^{t+1}$ 去求 $\mathbf{W}^{t+1}$ 。循环迭代就可得到最终解。

具体将 ICE 的优化问题简化为

$$\begin{cases} \min_{\mathbf{A} \in \mathbf{R}^{L \times M}, \mathbf{W} \in \mathbf{R}^{M \times N}} & \|\mathbf{Y} - \mathbf{AW}\|_F^2 + \lambda \text{Tr}(\mathbf{AHA}^T) \\ \text{s. t.} & \mathbf{W} \geq 0 \\ & \mathbf{W}^T \mathbf{L} = \mathbf{L} \end{cases} \quad (3.5.22)$$

将 $\mathbf{E} = \mathbf{Y} - \mathbf{AW}$ 代入式(3.5.22)中,并且使用 $(\mathbf{H} = \mathbf{I} - \mathbf{L}^T \mathbf{L} / \mathbf{M})$

$$\sum_{i \neq j} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 = \text{Tr}(\mathbf{AHA}^T) \quad (3.5.23)$$

另定义

$$J(\mathbf{A}, \mathbf{W}) = \|\mathbf{Y} - \mathbf{AW}\|_2^2 + \lambda \text{Tr}(\mathbf{A}^T \mathbf{H} \mathbf{A}) \quad (3.5.24)$$

如果在第 t 步迭代已得到 $\mathbf{A}^t$ 和 $\mathbf{W}^t$, 则 $\mathbf{W}^{t+1}$ 可由(固定 $\mathbf{A}$)求解下式得到:

$$\begin{cases} \mathbf{W}^{t+1} = \underset{\mathbf{W}}{\text{argmin}} J(\mathbf{A}^t, \mathbf{W}) \\ \text{s. t.} & \mathbf{W} \geq 0 \\ & \mathbf{W}^T \mathbf{L} = \mathbf{L} \end{cases} \quad (3.5.25)$$

式(3.5.25)是一个带线性限制条件的二次规划问题(QP),有标准的算法。接下来,固定 $\mathbf{W}$ 以得到下一步的原子信号矩阵 $\mathbf{A}$:

$$\mathbf{A}^{t+1} = \underset{\mathbf{A}}{\text{argmin}} J(\mathbf{A}, \mathbf{W}^{t+1}) \quad (3.5.26)$$

式(3.5.26)也是一个二次规划问题,且无约束条件,所以有直接的解析解:

$$\mathbf{A}^{t+1} = \mathbf{Y} (\mathbf{W}^{t+1})^T [\mathbf{W}^{t+1} (\mathbf{W}^{t+1})^T + \lambda \mathbf{H}]^{-1} \quad (3.5.27)$$

一般情况下,矩阵 $\mathbf{W}^{t+1} (\mathbf{W}^{t+1})^T + \lambda \mathbf{H}$ 是可逆的,因为 $\mathbf{H}$ 的秩为 $M-1$ 。只有在某些情况如 $\mathbf{W}^{t+1}$ 为零矩阵时才会出现 $\mathbf{W}^{t+1} (\mathbf{W}^{t+1})^T + \lambda \mathbf{H}$ 不可逆。即便在这种情况下,也可以引入一个小的扰动,即 $\mathbf{W}^{t+1} (\mathbf{W}^{t+1})^T + \lambda \mathbf{H} + \epsilon \mathbf{I}$ ($\epsilon \geq 0$, 但接近 0)取代原矩阵 $\mathbf{W}^{t+1} (\mathbf{W}^{t+1})^T + \lambda \mathbf{H}$ 以确保其可逆。所以,从一个最初的原子矩阵 $\mathbf{A}^0$ 出发,通过上述迭代过程,可以得到一组局部最优解。但该优化算法也存在类似非负矩阵分解优化算法的问题,ICE 的优化目标同样是非凸性的,所以上述算法虽然可以保证收敛,但也可能收敛到局部最优解,并不能确保全局最优。另外,解还依赖于初始值,不同的初始值可能会导致不同的解。

类似于对非负矩阵分解的改进(参见式(3.5.17)),也可以对式(3.5.21)进行相应的变形,以克服 ICE 的一些缺点或引入 ICE 中没有的性质。例如,可得到稀疏迭代约束端元化(SPICE):

$$\begin{cases} \min_{\mathbf{A} \in \mathbf{R}^{L \times M}, \mathbf{W} \in \mathbf{R}^{M \times N}} & \|\mathbf{E}\|_F^2 + \lambda_1 \sum_{i \neq j} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 + \lambda_2 \|\mathbf{W}\|_1 \\ \text{s. t.} & \mathbf{Y} = \mathbf{AW} + \mathbf{E} \\ & \mathbf{W} = \mathbf{AW} \\ & \mathbf{W}^T \mathbf{L} = \mathbf{L} \end{cases} \quad (3.5.28)$$

其中引入的是权重矩阵的稀疏性先验条件。

3.5.4 稀疏编码

稀疏编码就是寻找数据的一种字典表达方式,要使得每个数据都可以表示成为少数几个原子信号的线性组合。直观地说,就是在字典学习中对权重矩阵引入稀疏性约束条件。在稀疏编码中权重矩阵称为稀疏表达矩阵,借助前面的优化问题表示方法,稀疏编码的模型和优化目标可写成

$$\begin{cases} \min_{A \in \mathbf{R}^{L \times M}, W \in \mathbf{R}^{M \times N}} \|E\|_p \\ \text{s. t. } Y = AW + E \\ \|w_i\|_0 \leq K \quad i = 1, \dots, N \end{cases} \quad (3.5.29)$$

式(3.5.29)所表达的含义是找到用于表示数据 Y 的字典,使得使用该字典的每个数据的重建系数在一定误差允许范围内是 K 稀疏的。式(3.5.29)与压缩感知的经典优化问题有些相似之处,但有两点不同需指出:一是压缩感知中的讨论针对一个信号的感知与重建;二是压缩感知中的采样矩阵 Φ 对应稀疏编码中的矩阵 A ,且 Φ 已知,但在稀疏编码中矩阵 A 也是需要求解的目标之一。可见,稀疏编码在优化问题上求解的复杂度远超过压缩感知。

1. 最优方向法

最优方向法(MOD)解决的是式(3.5.29)及以下稍许不同的目标:

$$\begin{cases} \min_{A \in \mathbf{R}^{L \times M}, W \in \mathbf{R}^{M \times N}} \|W\|_0 \\ \text{s. t. } Y = AW + E \\ \|e_i\|_2 \leq \epsilon \quad i = 1, \dots, N \end{cases} \quad (3.5.30)$$

其中, $\|W\|_0$ 为矩阵 W 的 L_0 范数,即矩阵 W 中非零元素的个数。

最优方向法的本质是交替优化原子信号矩阵和权重矩阵,主要步骤有两个:

(1) 稀疏编码阶段: 对 $i = 1, \dots, N$, 使用匹配追踪算法以获取

$$w_i = \arg\min_A \|y_i - A^T w\|_2 \quad \text{s. t. } \|w\|_0 \leq K \quad (3.5.31)$$

(2) 字典更新阶段: 使用下面的公式更新原子信号矩阵

$$A^t = \arg\min_A \|Y - AW^t\|_F^2 = YW^t [W^t (W^t)^T]^{-1} \quad (3.5.32)$$

2. 核奇异值分解

在 MOD 中,字典更新是通过对权重矩阵进行简单的伪逆计算来实现的,这样会使一些原子信号混入到其他原子信号中并产生影响,从而导致整个算法收敛过程较慢。**核奇异值分解(K-SVD)**是为了消除原子信号之间的互相干扰而提出的。它的基本思想是逐个考查原子信号对数据拟合的贡献并逐个更新原子信号。

考虑第 j 个原子信号 a_j , 令 S_j 为包含原子信号 a_j 中数据的索引集合,即

$$S_j = \{i \mid 1 \leq i \leq M, w_{ji} \neq 0\} \quad (3.5.33)$$

其中, w_{ji} 为权重矩阵 W 的第 j 行第 i 列元素。

因为 W 是稀疏的,所以并不是每个原子信号都对观测数据有贡献,或者说并不是每个数据都包含所有原子信号,而是最多只有 K 个原子信号具有非零的权重系数。先来计算残差矩阵:

$$E_j = Y - \sum_{l \neq j} a_l w_l^T \quad (3.5.34)$$

式(3.5.34)计算的是除 a_j 外其他原子信号的拟合残差。引入记号 E_j^S 以表示矩阵 E_j 的列子矩阵,即由 E_j 中属于 S 的列所组成的矩阵。对 a_j 和 w_j 的更新由求解下式得到:

$$(a_j, w_j) = \arg\min_{a, w} \|E_j^S - aw^T\|_F^2 \quad (3.5.35)$$

式(3.5.35)仍是一个矩阵分解问题,但比原始的矩阵分解问题简单得多,因为该式中的 a 和 w 都是矢量而非矩阵。所以,对式(3.5.35)可用经典的矩阵分解来求解。在 K -SVD 中

使用对矩阵 E_j^S 的秩为 1 的 SVD 分解,即 tSVD

$$E_j^S \approx \mathbf{u} \lambda \mathbf{v}^T \quad (3.5.36)$$

可见, $\mathbf{u}$ 和 $\lambda \mathbf{v}$ 分别是 $\mathbf{a}$ 和 $\mathbf{w}$ 的解。如果 $K=1$ 且限定权重矩阵的值只能取 1 或 0,那么原始问题退化成一个标准的 K -means 聚类问题。类似于在每个 K -means 的迭代步骤中都计算 K 个子集(类),这种带前述限制条件的 K -SVD 的每个迭代步骤也是在寻找 K 个子矩阵(对应 K 个子集)。它的两个主要迭代步骤为

(1) 稀疏编码阶段: 对 $i=1, \dots, N$, 使用匹配追踪算法以获取

$$\mathbf{w}_i = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{y}_i - \mathbf{A}^{t-1} \mathbf{w}\|_2^2 \quad \text{s. t. } \|\mathbf{w}\|_0 \leq K \quad (3.5.37)$$

的近似解,这样得到 $\mathbf{W}^t$ 。

(2) 字典更新阶段: 使用下面公式依次更新原子信号矩阵 $\mathbf{A}$ 的第 j 个原子信号 $\mathbf{a}_j (j=1, \dots, M)$ 。

先计算残差矩阵

$$E_j = \mathbf{Y} - \sum_{l \neq j} \mathbf{a}_l \mathbf{w}_l^T \quad (3.5.38)$$

而矩阵 E_j^S 的秩为 1 的 tSVD 分解为

$$E_j^S \approx \mathbf{u} \lambda \mathbf{v}^T \quad (3.5.39)$$

$\mathbf{a}_j = \mathbf{u}$, 且权重向量 $\mathbf{w}_j = \lambda \mathbf{v}$ 。

可见,稀疏编码与聚类分析有很大的相似性,且与子空间分析(也称子空间学习)密切相关。

3.6 压缩感知的成像应用

压缩感知已得到了许多应用,下面仅举两个成像方面的例子。

3.6.1 单像素相机

单像素相机的设计是使压缩感知得到广泛关注的一个重要事件。它成像灵活,光电转换效率高,大大减少了对高复杂度、高代价光探测器的要求。它的成像流程如图 3.6.1 所示。



图 3.6.1 单像素相机成像流程

利用光学透镜将场景中得到光源照射的目标投影到数字微镜器(DMD)上。这是一种借助大量的微小镜片来反射入射光而实现光调制的器件。微镜阵列中的每个单元都可借助电压信号控制以分别进行 $\pm 12^\circ$ 的机械翻转,如此就可将入射光分别进行对称角度的反射或完全吸收而不输出。这样就构成了一个由 1 和 0 组成的随机测量矩阵。以对称角度反射出来的光被光敏二极管(目前常用的快速、敏感、低价高效的单像素传感器,在低光照时也有使用血崩二极管的)所接收,其电压随反射光强度发生变化。量化后可给出一个测量值 y_i ,其中每次 DMD 的随机测量模式对应测量矩阵中的一行 ϕ_i ,此时如果将输入图像看成一个矢

量 x , 则该次测量的结果为 $y_i = \Phi_i x$ 。将此投影操作重复 M 次, 则通过 M 次随机配置 DMD 上每个微镜的翻转角度, 就可获得 M 个测量结果, 得到 $y = \Phi x$ 。根据总变分重构法(可用 DSP 实现), 就可用远小于原始场景图像像素的 M 个测量值重构图像(其分辨率与微镜阵列的分辨率相同)。这相当于在图像采集的过程中实现了对图像数据的压缩。

例 3.6.1 单像素相机成像效果

单像素相机成像的效果目前还离普通 CCD 相机成像的效果有一定的距离。图 3.6.2 给出一组示例。图(a)是对一张黑纸上的白色字母 R 的成像效果, 像素个数为 256×256 ; 图(b)是一个用单像素相机成像的效果, 此时 M 为像素个数的 16%(进行了 11 000 次测量); 图(c)是另一个用单像素相机成像的效果, 此时 M 为像素个数的 2%(进行了 1300 次测量)。

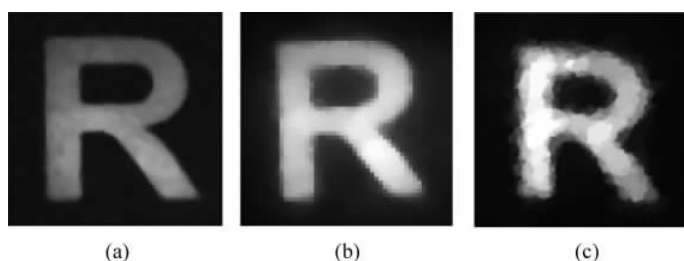


图 3.6.2 单像素相机成像效果示例

由图可见, 虽然测量数量有所减少, 但质量还不能相比。另外, 机械翻转微镜需要一定的时间, 所以为获得足够多的测量数量的成像时间比较长(以分钟为单位)。事实上, 在可见光范围内, 用单像素相机成像的成本比一般 CCD 相机或 CMOS 相机要高。由于可见光谱与硅材料的光电响应区域一致, 所以 CCD 或 CMOS 器件的集成度高且价格低。但在其他光谱范围, 例如红外谱段, 单像素相机就比较有优势。由于红外谱段的探测器件很昂贵, 单像素相机就可与其竞争了。□

3.6.2 压缩感知磁共振成像

磁共振成像是一种借助扫描并进行投影重建来获取图像的方法(参见上册第 8 章), 在医学中得到广泛应用。磁共振成像可获取人体内部的信息, 且对人体无损伤, 但其中的扫描过程导致采集图像需要大量的时间, 不仅增加了病人的不适感, 而且病人的生理运动造成的运动伪影也会给临床诊断带来较大的负面效应。因此, 缩短成像时间、提高成像效率和减少图像伪影得到很多关注。

利用压缩感知的磁共振成像的基本思路: 首先对图像采用离散傅里叶标准正交基进行稀疏表达, 其次对得到的 K 空间数据进行随机欠采样, 最后通过非线性重构算法重建出图像。整个流程如图 3.6.3 所示。

具体描述如下:

(1) 考虑一幅 N -D 图像 x 在 $\Phi = \{\Phi_1, \dots, \Phi_N\}$ 下具有稀疏性, 则

$$x = \sum_{i=1}^K \theta_i \Phi_i = \Phi \theta \quad (3.6.1)$$

其中, $\theta = [\theta_1, \dots, \theta_N]^T (K \ll N)$ 为 x 在稀疏变换 Φ 中的非零系数。

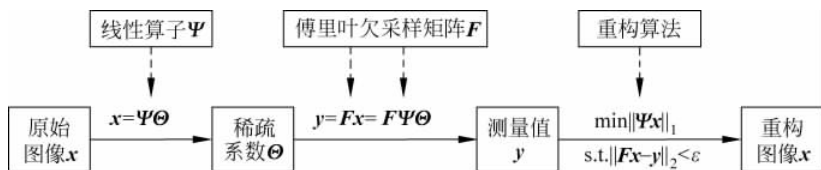


图 3.6.3 压缩感知磁共振成像流程

(2) 借助 $F \in \mathbf{R}^{M \times N}$ ($M \ll N$) 将图像 x 投影到低维空间, 得到 M 维的测量值 $y \in \mathbf{R}^M$:

$$y = Fx = F\Psi\Theta \quad (3.6.2)$$

其中, F 为傅里叶变换欠采样矩阵。

(3) 图像 x 可由 K 空间的测量值 y 通过求解约束优化问题来精确重构:

$$\min \|\Psi x\|_1 \quad \text{s.t.} \quad \|Fx - y\|_2 < \epsilon \quad (3.6.3)$$

式(3.6.3)表示在用不等式约束 K 空间数据的一致性条件下(ϵ 为测量值重构的保真度)通过稀疏变换 Ψ 对 x 进行稀疏表达。

例 3.6.2 压缩感知磁共振成像效果

图 3.6.4 给出一些压缩感知磁共振成像的效果示例。图像为 T_1 加权脑图像, 每幅图像的分辨率为 384×324 , 含有加性复数高斯噪声, 信噪比为 25dB。其中, 左边 3 幅图是采用基于压缩感知的总变分重构法得到的结果, 右边 3 幅图是采用基于压缩感知的凸优化重构法(使用了加权 L_1 范数)得到的结果。每 3 幅图中, 从左向右, 测量数分别是原始像素数的 $1/4$, $1/6$ 和 $1/8$ (都对应欠采样)。换句话说, 成像的加速倍率(反比于测量数)分别为 4, 6, 8。

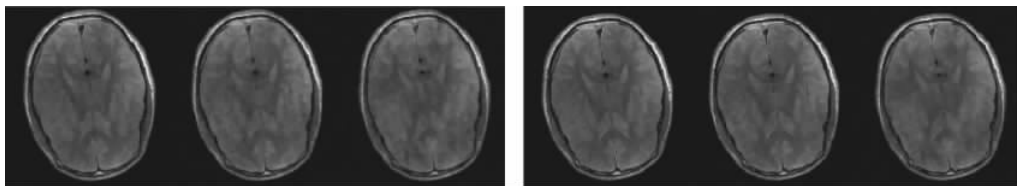


图 3.6.4 压缩感知磁共振成像效果示例

由图 3.6.4 可见, 当加速倍率为 6 和 8 时, 利用总变分重构法得到的结果中解剖结构比较模糊且存在较多块状伪影。利用凸优化重构法得到的结果有所改善, 但上述问题仍然存在。□

在利用压缩感知的磁共振成像中, 扫描仪器采集的并不是原始图像像素, 而是由全局傅里叶变换得到的频域图像, 每一个频域像素都是时域像素的线性组合, 即频域图像包含了原始图像的所有信息。因此, 保留部分重要的采集数据并不会导致原始图像信息的缺失。然后, 通过仅利用 K 空间的部分数据就可成功地重构出原始图像。如此, 不但加快了成像速度, 而且将极大地减少扫描时间并提高病人的舒适性。

总结和复习

为更好地学习, 下面对各小节给予概括小结并提供一些进一步的参考资料; 另外给出一些思考题和练习题以帮助复习(文后对加星号的题目还提供了解答)。

1. 各节小结和文献介绍

3.1 节概括给出了压缩感知的基本框架,相关的内容还可参见[卢 2012]、[万 2013]、[尹 2013]、[荆 2015]、[闫 2015]等。

3.2 节介绍了稀疏表达的一些基本知识,可与 3.5 节结合学习。有关稀疏表达的一个应用可见[Liu 2014]。

3.3 节分析了测量矩阵及其特性,相关讨论在很多书籍中都有介绍,如[李 2015]。

3.4 节介绍了一些典型的重构算法,凸优化算法还包括:① L_p 重构模型,用 L_p ($0 < p < 1$) 范数代替 L_0 范数;②平滑 L_0 重构模型;③低秩重构模型,利用了图像低秩结构;④非局部重构模型,利用了图像自相关结构;⑤结构化稀疏聚类模型,结合了局部和非局部模型。贪婪算法的种类也有很多,其中,匹配追踪(MP)算法、正交匹配追踪(OMP)算法能够以极大的概率重构原信号,但是缺乏相应重构正确性的保证。随后改进的算法有逐步匹配追踪(StOMP)、子空间追踪(SP)、压缩匹配追踪(CoOMP)和正则匹配追踪(ROMP)等,在一定程度上兼顾重构的通用性和复杂性,并提供了理论支撑。对于低维度的小尺度信号重构,贪婪算法的速度快,但是对于含噪声的大尺度重构问题,该算法不具有鲁棒性。对压缩感知中图像重构的模型综述可见[宋 2016]。

3.5 节概括地对稀疏编码与字典学习进行了介绍。稀疏编码还有许多拓展,即为了获取不同性质的稀疏编码的变形算法[李 2015]。如双稀疏模型(double-sparsity model),正交单位基组成的字典(union of unitary bases),图像标记字典(image signature dictionary, ISD)等。关于非负矩阵近期研究进展的一个综述可见[Wang 2013]。关于字典学习的一个应用可见[Liu 2016]。一种融合解析模型和综合模型的压缩感知算法可见[练 2016]。

3.6 节只列举了两个压缩感知在成像上的应用示例。基于压缩感知理论的研究工作很多,有关磁共振血管快速成像技术(CS-MRA)的进展可见[赵 2016],有关雷达成像技术与应用的进展可见[李 2016]。更多应用还可见[王 2014]。

2. 思考题和练习题

3-1 有人将压缩感知的基本流程图用图题 3-1 表示,讨论其与图 3.1.1 的相似和不同处。



图题 3-1

3-2 什么样的信号可认为是稀疏信号? 举 3 个实例。

*3-3 对比传统的信息获取和压缩感知信息获取。

(1) 在传统的信息获取中压缩的是数据,而在压缩感知中,压缩的是什么?

(2) 在传统的信息获取中只需要解压缩,而在压缩感知中需要重构步骤,为什么?

3-4 试画出范数 $3/2$ 与拟范数 $1/4$ 在 $\mathbf{R}^2$ 中的单位圆。

3-5 试讨论使用范数 $3/2$ 与拟范数 $1/4$ 在描述逼近误差时的不同表现。

3-6 考虑对周期信号 $y = \sin(0.25\pi i)$ 在 $i = 0, \dots, 16$ 进行采样得到的信号,

- (1) 画出对其进行傅里叶变换的频谱；
- (2) 画出对其进行小波变换的频谱；
- (3) 比较两个频谱,哪个更加稀疏? 为什么?

3-7 试证明: 如果 (Φ, Δ) 满足式(3.3.4), 则矩阵 Φ 满足 $2K$ 阶零空间特性。

3-8 对比列出文中介绍的各类重构算法的适用场合、优缺点。还可扩展到其他文献中的算法。

3-9 试证明式(3.4.7)的解 $\mathbf{x}'$ 满足式(3.4.8)。

3-10 非负矩阵分解对原子信号和权重矩阵有什么限定? 试举例解释其物理意义。

***3-11** 压缩感知的目标和稀疏编码的目标有什么不同? 哪一个的要求更高些?

3-12 调查一下, 压缩感知除在成像方面外, 还在哪些领域得到了应用, 介绍几个示例。