



2006年,Donoho和Candès等提出了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中可压缩或稀疏时,通过采集少量信号的非自适应线性投影值,建立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利用远少于奈奎斯特-香农采样数量的线性、非自适应测量值无失真地重建出来。

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3方面:用于稀疏表达(表示)的稀疏编码、用于测量编码的测量矩阵及用于重建信号的重构算法。其中,稀疏表达是压缩感知理论的先验/先决条件。测量编码是压缩感知硬件实现的关键,但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。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,将信号变换到一个新的基或框架下,当其非零系数的个数远少于原始信号的项数时,即可实现压缩感知。

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

3.4节介绍基本的重构模型,分析无噪声和有噪声时重构稀疏信号的原理,并列举几种典型的重构算法,包括基于深度学习的重构算法。

3.5节讨论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,而字典学习的目标是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如图3.1.1所示,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(每个方框的左边为输入,右边为输出),第二行的3个方框分别指示实现压缩感知要完成的3项工作(分别对应3个关键处理步骤)。

对压缩感知的3个关键处理步骤进一步说明如下。

(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,真正稀疏的自然信号很少。某个信号稀疏严格地说是指该信号可通过稀疏信号近似表达。主要工作是找



图 3.1.1 压缩感知的整体流程框图

到对信号可压缩的某个正交基或紧框架(见下),从而实现对信号的稀疏表达。从数学角度看,是指构建稀疏变换矩阵,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成功实现的关键。对观测矩阵的设计不仅关系到压缩和采样过程的快慢,而且要保证对稀疏矢量降维时重要信息不受破坏,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)精确重构高维原始信号。主要工作是设计快速重构算法,从数量较少的线性观测中恢复信号。重构算法的性能决定了恢复信号的质量。

例 3.1.1 压缩感知采样与奈奎斯特-香农采样的对比

从信号处理的角度看,压缩感知采样与奈奎斯特-香农采样有很多区别。图 3.1.2 为其基本流程对比,左边为香农采样的流程,高速采样后进行压缩以减少数据量,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,在需要原始信息时进行图像重构。

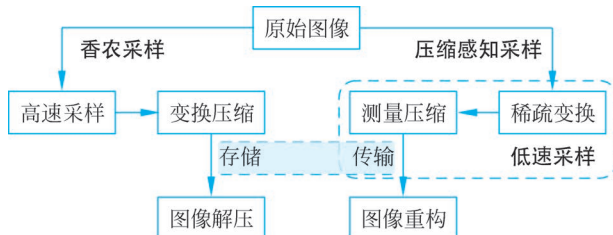


图 3.1.2 压缩感知采样与奈奎斯特-香农采样的流程对比

表 3.1.1 列出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 压缩感知采样与奈奎斯特-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以获取采样	非直接采样: 通过采样/测量矩阵与目标信号相乘间接地获取采样/测量值
采样特点	均匀采样: 用目标信号中最高频率两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两倍, 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数非零信号, 则只需较少的采样/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 函数插值即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

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,但真正稀疏,能直接满足压缩感知要求的信号很少。一般认为某个信号稀疏,是指这个信号可以通过稀疏信号近似地表达。下面具体讨论。

先复习、介绍和讨论一些预备知识。

1. 矢量空间

下列讨论中将有限域中的离散信号看作分布在 N -D 欧氏空间中的矢量,将此空间记为 $\mathbb{R}^N$ 。在 $\mathbb{R}^N$ 中,范数是一个重要的描述概念。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, $|S|$ 表示集合 S 的基数,即集合 S 中元素的个数。

常用的范数包括 L_1 、 L_2 和 L_∞ , 其特性分别如图 3.2.1(a)、图 3.2.1(b)、图 3.2.1(c) 所示(可对比上册图 3.1.5)。在 $p < 1$ 的情况下,式(3.2.1)中定义的范数已无法满足三角不等式,所以其本质上为拟范数。拟范数对应的 $\mathbb{R}^2$ 中的单位圆不再是凸集。例如,拟范数 $L_{1/2}$ 对应的 $\mathbb{R}^2$ 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 对应的 $\mathbb{R}^2$ 中的单位圆 $\{\mathbf{x}: \|\mathbf{x}\|_0 = 1\}$ 如图 3.2.1(e) 所示。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。

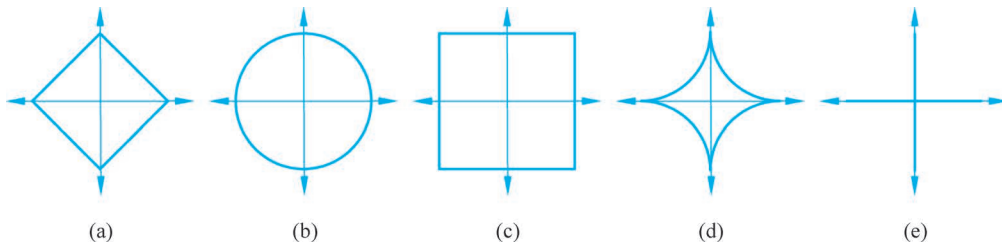


图 3.2.1 不同范数在 $\mathbb{R}^2$ 中的单位圆

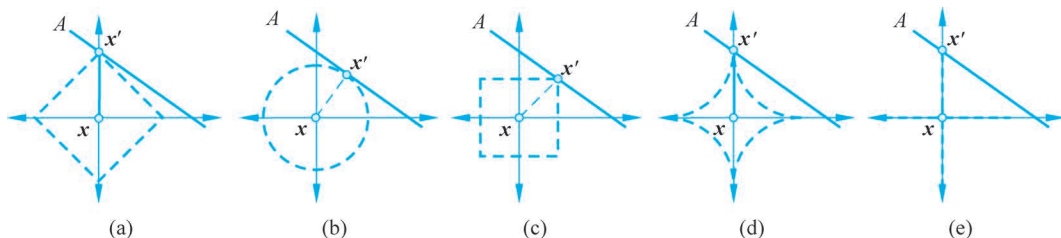
例 3.2.1 不同范数在描述逼近误差时的不同表现

范数可用于描述信号的强度或误差的大小。假设已知一个信号 $\mathbf{x} \in \mathbb{R}^2$, 希望用 1-D 仿射空间 A 中的点逼近它。若采用 L_p 范数衡量这种逼近误差, 则任务是找到 $\mathbf{x}' \in A$, 使 $\|\mathbf{x} - \mathbf{x}'\|_p$ 最小。对参数 p 的选择很关键, 不同的 p 值将使逼近误差呈现不同的特性表现。为了找出 A 中最接近 $\mathbf{x}$ 的点, 可以想象一个以 $\mathbf{x}$ 为中心的 L_p 球不断膨胀, 碰到 A 的点即为 L_p 条件下最接近 $\mathbf{x}$ 的点 $\mathbf{x}'$ 。如图 3.2.2 所示, 图 3.2.2(a) 对应 L_1 范数, 图 3.2.2(b) 对应 L_2 范数, 图 3.2.2(c) 对应 L_∞ 范数, 图 3.2.2(d) 对应 $L_{1/2}$ 拟范数, 图 3.2.2(e) 对应 L_0 拟范数。

由图 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, 误差的衡量方式将有很大概率使选择的 $\mathbf{x}'$ 与 $\mathbf{x}$ 位于同一水平或垂直轴上, 即非对称地将误差缩小到 1-D 空间中(减少了一个维度), 从而促进稀疏特性产生。□

2. 基和框架

考虑矢量集 $\Psi = \{\phi_i\}, i = 1, 2, \dots, N$, 如果其中的矢量可以生成矢量空间 $V = \mathbb{R}^N$, 且矢量

图 3.2.2 5 种 p 值的 L_p 描述逼近误差时的不同表现

ϕ_i 之间是非线性相关的, 则 Ψ 可称为有限维矢量空间 V 中的基。换句话说, 矢量空间 V 中的任一矢量都可通过 Ψ 中矢量的线性组合唯一地表达, 且线性表达的系数可通过使用信号和该矢量空间中基的内积表示。

表示矢量集的符号也可用于表示矩阵, 如 Ψ 可表示由列矢量 ϕ_i 构成的矩阵, 其大小为 $N \times N$ 。标准正交集是一种特殊的基, 定义为: 矢量集 $\Psi = \{\phi_i\}$, 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 $\Psi^T \Psi = I$, I 表示 $N \times N$ 的单位矩阵。此时, 对于该矢量空间中的任意矢量 x , 可以容易地计算出该标准正交集表示下的系数 a , 即 $a = \Psi^T x$ 。

将基的概念推广到可能线性相关的矢量集, 就形成了列框架, 即矢量集 $\Psi = \{\phi_i\}_{i=1}^N$, 且 $\phi_i \in \mathbb{R}^d$, 其中 $d < N$, 相当于矩阵 $\Psi \in \mathbb{R}^{d \times N}$, 对于所有矢量 $x \in \mathbb{R}^d$ 满足

$$L \|x\|_2^2 \leq \|\Psi^T x\|_2^2 \leq R \|x\|_2^2 \quad (3.2.2)$$

其中, $0 < L \leq R < \infty$ 。

注意, $L > 0$ 表示矩阵 Ψ 中的行矢量一定是线性独立的。当 L 为使不等式成立的最大值, R 为使不等式成立的最小值时, 则将其称为无框架界。如果 $L = R$, 则称此框架为紧框架。如果 Ψ 为有限维矩阵, 则 L 和 R 分别对应 $\Psi \Psi^T$ 的最小特征值和最大特征值。由于框架具有一定的冗余性, 所以其可为目标数据提供更丰富的表达, 即针对一个目标信号矢量 x , 存在无数个系数矢量 a , 使 $x = \Psi a$ 。

3. 稀疏性表达

为了更精练地表达信号, 可以将信号变换到新的基或框架下, 当非零系数的个数远少于原始信号的项数时, 可以将这些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。只要预先确认目标信号在已知基或框架下具有稀疏性表达, 就可无失真地重建原始信号。在稀疏性表达中, 常将基或框架称为字典或过完备字典, 而将其中的矢量元素称为原子。

从数学角度看, 当信号 x 中最多有 K 个非零值时, 称信号 x 是 K 稀疏的, 即 $\|x\|_0 \leq K$, 可采用 $\Sigma_K = \{x : \|x\|_0 \leq K\}$ 表示所有 K 稀疏信号的集合。另外, 对于一些本身并不稀疏但在一些基矩阵 Ψ 中具有稀疏性表达的信号, 只要 $x = \Psi a$, 其中 $\|a\|_0 \leq K$, 则仍可将这些信号看作 K 稀疏的。

3.3 测量矩阵及特性

下面先比较传统方法和压缩感知方法是如何获取信号的, 再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, 对于一个只包含 K 个非零值的信号 $x \in \mathbb{R}^N$, 需采集

N 个值,才能保证完全保持 x 中的信息。具体采样过程可用采样矩阵 Φ 与信号矢量 x 的乘积表示,即

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

采样过程可用图 3.3.1 表示。此时采样矩阵对应角单位矩阵,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,必然带来较大的浪费。

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。此过程仍可用式(3.3.1)的形式表示,但其中 Φ 为固定的 $M \times N$ 矩阵,采样所得的测量值 $y \in \mathbb{R}^M$ 。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 对应降维投影操作,即把 $\mathbb{R}^N$ 映射到 $\mathbb{R}^M$ 中。一般有 $K < M \ll N$,即矩阵 Φ 的列数远多于行数。

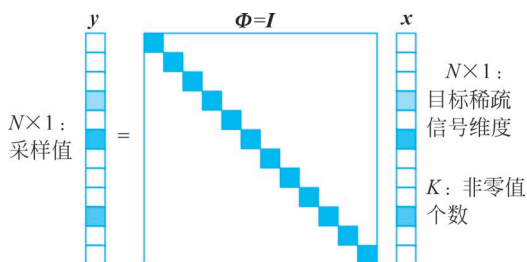


图 3.3.1 奈奎斯特-香农采样模型

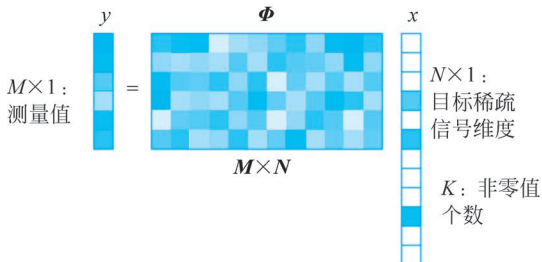


图 3.3.2 针对本身稀疏信号的测量模型

由前面对稀疏表达的讨论可知,对于本身并不稀疏的信号,可将其变换到新的基或框架下,使其表现出稀疏特性(可理解为通过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 $x = \Psi z$,则可将 Φ 与 Ψ 合并,而测量过程仍用式(3.3.1)的形式表示。不过这种推广的、具有普适性的测量模型用图 3.3.3 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。

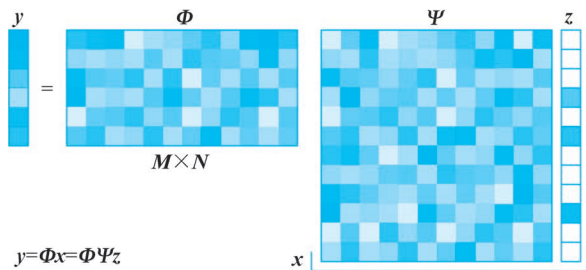


图 3.3.3 具有普适性的测量模型

3.3.2 测量矩阵特性

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

对于一个矩阵 $\Phi: \mathbb{R}^N \rightarrow \mathbb{R}^M$,其零空间 $\mathcal{N}(\Phi)$ 可定义为

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

对于任意稀疏信号 x ,若希望基于测量值 $y = \Phi x$ 无失真地重建该稀疏信号 x ,则对于任意两个矢量 $x, x' \in \Sigma_K = \{z: \|z\|_0 \leq K\}$,一定有 $\Phi x = \Phi x'$,否则基于测量值 y 无法区别 x 和

$\mathbf{x}'$ 。假设 $\Phi \mathbf{x} = \Phi \mathbf{x}'$, 则 $\Phi(\mathbf{x} - \mathbf{x}') = \mathbf{0}$, 其中 $\mathbf{x} - \mathbf{x}' = \mathbf{h} \in \Sigma_{2K}$ 。由此可见, 用矩阵 Φ 可唯一表达 $\mathbf{x}$ 的充分必要条件是: Φ 的零空间 $\mathcal{N}(\Phi)$ 中不含 Σ_{2K} 中的任何元素, 即 $\mathcal{N}(\Phi)$ 与 Σ_{2K} 的交集为空集。

当要处理的是近似稀疏信号(含有少数明显的非零元素, 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种信号)时, 需要对测量矩阵 Φ 的零空间引入更严格的条件, 以确保 $\mathcal{N}(\Phi)$ 中不包含稀疏的矢量, 也不包含可压缩的近似稀疏的矢量。

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, 且该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。若存在一个常数 $C > 0$, 使下式对所有 $\mathbf{h} \in \mathcal{N}(\Phi)$ 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。

现用 $A: \mathbb{R}^M \rightarrow \mathbb{R}^N$ 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 N 个原始目标信号的算法, 且 $M \ll N$ 。对于所有的 $\mathbf{x}$, 可要求该算法确保下式成立:

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

其中, $\sigma_K(\mathbf{x})_p = \min \|\mathbf{x} - \mathbf{x}'\|_p$, 表示去除矢量 $\mathbf{x}$ 的 K 个幅值最大元素后的 p 范数, 即若 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)可保证该重建算法能正确重建出所有可能的含 K 个非零元素的信号, 同时也能使利用重建的稀疏信号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, A) 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。当测量值有噪声或量化引入误差时, 需要讨论更严格的重建条件。为此, 定义 K 阶约束等距特性(RIP)如下。

若存在 $\delta_K \in (0, 1)$, 使

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

对于所有 $\mathbf{x} \in \Sigma_K \{\mathbf{x}: \|\mathbf{x}\|_0 \leq K\}$ 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 是对于所有 K 阶稀疏矢量 $\mathbf{x}$ 均满足上式的最小常数)。

矩阵的 RIP 是针对两个参数 δ_K 和 K 而言的。若某一个矩阵满足 RIP, 则对特定的 δ_K 和 K , 式(3.3.5)对任意 $\mathbf{x} \in \Sigma_K$ 均成立。若矩阵 Φ 满足 $2K$ 阶约束等距特性, 则可将式(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, 其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到重要作用。同时表明 K 阶稀疏的矢量不在测量矩阵的零空间中, 如果在其中, 信号就无法重建了。约束等距特性也可表示为: 从测量矩阵中随机选取 $2K$ 列, 构成的子集只能“近似正交”。因为子集的行列长度不同, 所以不可能真正正交。

假设利用测量矩阵 Φ 获取 1-D 的 K 阶稀疏矢量 $\mathbf{x}$ 的线性测量值 $\mathbf{y}$, 当 $\delta_{2K} < 1$ 时, 即可基于测量矢量 $\mathbf{y} = \Phi \mathbf{x}$ 重建出原始 K 阶稀疏矢量, 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, 若要重建出独一无二的 K 阶稀疏矢量, 必须确保 $\delta_{2K} < 1$, 同时约束等距常数 δ_K 不能为 0, 如果 $\delta_{2K} = 0$, 则说明测量矩阵 Φ 为标准正交矩阵, 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取决于测量矩阵或约束等距常数。但是,某测量矩阵是否满足这个特性,无法通过计算的方式判断,因为这是一个 NP 难题。

若一个矩阵满足约束等距特性(RIP),则其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 M 相对较大。如果在构建过程中引入随机性,可降低测量个数的要求。通过随机方法构建的测量矩阵也称通用测量矩阵。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即可实现利用最少采样个数完成采样的目的。

3.4 解码重构

压缩感知中的编码重构(也称重建)的目标是基于较少的线性观测值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,再介绍典型的重构算法。

3.4.1 重构原理

下面先介绍一种直观的重构模型,借此讨论重构的原理,并分析无噪声稀疏信号重构问题和有噪声稀疏信号重构问题。

1. 基本重构模型

假设矢量 $\mathbf{x} \in \mathbb{R}^N$ 是一个长度为 N 的稀疏信号,测量矩阵 $\Phi: \mathbb{R}^N \rightarrow \mathbb{R}^M$ 已知,现基于较少数量的测量值 $\mathbf{y} \in \mathbb{R}^M$ ($M < N$),重构原始目标信号 $\mathbf{x}$ 。

最直接的重构方法可描述为

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

若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

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

可见,设 $\Theta = \Phi \Psi$,则本质上式(3.4.1)与式(3.4.2)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 范数近似非凸的 L_0 范数,将组合优化问题转化为凸优化问题。换句话说,通过近似可将一个无法求解的问题转化为一个可通过现代优化理论求解的问题,在统计学中也称 Lasso。

当测量值无噪声时,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。 $B(\mathbf{y})$ 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 $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^N$, 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 范数最小化方法产生的误差建立了界限, 其中的前提条件是测量矩阵 Φ 满足 $2K$ 阶约束等距特性。针对具体的 $B(\mathbf{y})$ 构建出特殊的界限条件, 需考虑 $\mathbf{x}' \in B(\mathbf{y})$ 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 $O(K \ln(N/K))$ 个采样值就可无失真地重建任意包含 K 个非零元素的目标信号 $\mathbf{x}$, 而无须考虑 K 个非零元素如何分布。

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。 $B(\mathbf{y})$ 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

若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 $\mathbf{e}$ 是测量过程产生的误差。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 都为仅依赖于 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 $\mathbf{e} \in \mathbb{R}^M$ 的各分量来自均值为零、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 , 对于任何 $\epsilon > 0$ 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, 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

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

两种情况下的重构都是有界的, 表明压缩感知确实可应用于实际。

3.4.2 测量矩阵的校准

在讨论具体重构算法前, 还要分析一个问题。如果测量矩阵本身存在噪声, 还要考虑测量矩阵的校准问题。

回到式(3.4.3)和式(3.4.4),测量矩阵可能并不是确切已知的,它可能通过模型描述,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,随着客观工作条件的变化,测量矩阵的物理条件也可能漂移。为解决以上问题,常用以下4种方案。

(1) 忽略这些问题,但这可能明显影响重建精度。

(2) 简单地将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 $y = \Phi'x + \epsilon$,将压缩感知的重建问题转换为求解下列问题:

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

其中, η 表示测量矩阵不精确产生的误差。实际中很难获得 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 $y_l = \Phi'_l x_l + \epsilon_l$,将所有矢量排列起来组成矩阵的形式,有

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

其中,需要基于已知的训练信号和测量值校准测量矩阵:

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

即要优化 Φ' ,使 $\|\mathbf{Y} - \Phi'X\|$ 最小。

这种方案在能够获得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则无法使用。如果可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 $a = [a_1 \ \dots \ a_j \ \dots]$ 。监督校准问题就可转化为下列凸优化问题:

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

(4) 非监督校准:针对少数未知信号(需要保证具有稀疏的特性),利用其各自有限的、基于压缩感知模型的采样值,开展盲校准,即无须采用已知训练信号校准测量矩阵,或训练的目标稀疏信号是未知的。具体是将未知的训练信号矢量 $x_1, x_2, \dots, x_l$ 表述为矩阵 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 ,通过一定的方法确定增益矩阵 D 和 X ,即解决下列问题:

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

不过,求解式(3.4.15)可能导致出现无意义的解。如果对矩阵 D 和 X 没有限制,那么只要将矩阵 D 设为无穷大、使矩阵 X 趋于零,即可满足式(3.4.15),但这样的解没有实际意义。

可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 δ_i 排列在对角线上形成的对角矩阵。为避免前述无意义解 $\Delta = X = 0$,引入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 $D = \operatorname{diag}(d_i)$, $\sum_i d_i^{-1} = M$ 。这个约束条件也可替换为 $\delta_i = 1$ 等其他形式。由于可对 Δ 引入任意约束, Δ 和 D 本质上存在比例关系。以非监督方式校准测量矩阵的方法可表示为

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

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。因为不同的应用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。目前存在许多重构算法,其在运算时间、重构精度和稳定性等方面各有特点。

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: 要求测量值包含噪声或测量系统本身具有系统噪声时,重构算法也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的重构。

(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。

考虑上述部分或全部因素而得到的多数重建算法可分为 4 大类: 凸优化算法、贪婪算法、组合算法和贝叶斯算法。下面分别针对每大类介绍一两个典型算法。

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 $\mathbb{R}^N$ 空间的凸集中优化关于未知变量 $\mathbf{x}$ 的凸目标函数 $J(\mathbf{x})$ 。这类算法重构精度高,需要的测量数据较少,约为 $O(K \log(N/K))$; 但计算速度慢,计算复杂度约为 $O(N^3)$ 。

假设 $J(\mathbf{x})$ 是一个能促进稀疏性(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。当测量值无噪声时,基于测量值 $\mathbf{y} = \Phi \mathbf{x}$ 要重建信号 $\mathbf{x}$,可解下列方程:

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

而当测量值有噪声时,可解下列方程:

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

其中, H 是对 $\Phi \mathbf{x}$ 和 $\mathbf{y}$ 之间矢量距离进行惩罚的代价函数。

式(3.4.18)也可改写为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 是一个惩罚因子,它可以通过试错法或交叉验证法选取。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 $\mathbf{y}$ 与理论测量值 $\Phi \mathbf{x}$ 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 $H(\Phi \mathbf{x}, \mathbf{y})$, 被称为 Lasso 问题。

(1) 总变分重构法

从广义角度看, $J(\cdot)$ 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,又在已知变换域 Ψ 中体现出稀疏性,则可以有如下混合正则项:

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

其中, $\text{TV}(\mathbf{x})$ 是目标信号的总变分(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 为二次型(凸函数且可导),则常使用收缩循环迭代法,也称软阈值法求解。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

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

用收缩循环迭代法也可解决式(3.4.19)的问题,基本方法可写为定点迭代的方式:对于 $i=1, 2, \dots, N$, 可将目标信号 $\mathbf{x}$ 的第 i 个分量的第 $k+1$ 次迭代步骤写为

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

其中, $\tau > 0$ 是梯度下降法的步长,可随着 k 而变化; μ 可根据噪声大小和经验选取, μ 越大表明允许 $\mathbf{x}_i^{k+1}$ 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,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:

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

针对这种特殊的惩罚函数,式(3.4.23)可简化为

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

3. 贪婪算法

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

$$\min \{ |\mathbf{I}| : \mathbf{y} = \sum_{i \in \mathbf{I}} \phi_i x_i \} \quad (3.4.26)$$

其中, $\mathbf{I} \subseteq \{1, 2, \dots, N\}$, 表示一个索引集(对应支撑集); ϕ_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需要的迭代次数决定的,计算速度一般较快,但需要的测量数据多且重构精度较低。

(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。匹配追踪将信号看作由基/字典中的元素通过线性组合而成。在压缩感知的稀疏重构中,字典就是采样矩阵 $\Phi \in \mathbb{R}^{M \times N}$,其每一列表示一种原型信号(也称原子信号)的元素。重构可描述为:给定一个信号 $\mathbf{y}$,寻求通过基/字典中元素的稀疏性组合描述信号 $\mathbf{y}$ 。

匹配追踪算法主要围绕原始观测信号 $\mathbf{y}$ 与线性组合的残差 $\mathbf{r} \in \mathbb{R}^M$ 展开,残差主要描述未被解释的测量值。每次迭代从字典中选取一个与残余分量差相关性最大的列:

$$\lambda_k = \underset{k}{\operatorname{argmin}} \left\{ \frac{\langle \mathbf{r}_{k-1}, \phi_{\lambda} \rangle \phi_{\lambda}}{\|\phi_{\lambda}\|^2} \right\} \quad (3.4.27)$$

其中, ϕ_{λ} 表示矩阵 Φ 的第 λ 列。一旦这列被选中,就获得了一个更逼近原始信号的结果。因为一个新的系数 λ_k 加入到对原始信号逼近的索引集中。

再进行如下更新:

$$\mathbf{r}_k = \mathbf{r}_{k-1} - \frac{\langle \mathbf{r}_{k-1}, \phi_{\lambda_k} \rangle \phi_{\lambda_k}}{\|\phi_{\lambda_k}\|^2}$$

$$\mathbf{x}^k = \mathbf{x}^{k-1} + \langle \mathbf{r}_{k-1}, \boldsymbol{\phi}_{\lambda_k} \rangle \quad (3.4.28)$$

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

上述算法存在两个缺点:一方面,不能保证重建误差足够小;另一方面,常需要较大的循环量才能逼近原始信号。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,每次循环都要将残差 $\mathbf{r}$ 投影到与所有选定列线性展开的正交子空间中,而不是用残差减去与其最大相关的字典中的矢量。由于残差 $\mathbf{r}$ 总是与已选取的列正交,所以相同的列在 OMP 中不会被选中两次,因而其最大循环次数会明显减少。

假设 Φ_k 表示矩阵 Φ 在第 k 个步骤中选出的子矩阵(第 k 列的选取过程与 MP 相同),则

$$\mathbf{x}_k = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \Phi_k \mathbf{x}\| \quad \mathbf{r}_k = \mathbf{y} - \Phi_k \mathbf{x}_k \quad (3.4.29)$$

可以证明,如果原始待测量信号是稀疏的,同时测量矩阵中每个元素都是从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。此算法最多需要 K 次循环即可收敛,其中 K 为原始信号中的非零个数。该算法可保证每次迭代的最优性,从而减少迭代次数。但其每次迭代仅选取一个元素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) 分段正交匹配追踪算法

分段正交匹配追踪(StOMP)是对 OMP 的一种改进。与 OMP 每次循环仅从字典中选取一个元素不同,StOMP 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 $O(MN \ln N)$,远低于 OMP。

(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进行精确的信号重建和逼近的信号重建。对于任意测量矩阵,只要其满足约束等距特性,SP 就能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,最重要的是确定测量矢量 $\mathbf{y}$ 位于哪个子空间,以及这些子空间由测量矩阵中哪 K 个列矢量生成。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即可计算出信号的非零系数。SP 的重要特点是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,检测这个子空间能否较好地重建信号。假如重建信号与实际信号之间的残差未达到算法的要求,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,同时加入相同数量的新候选值更新子空间。更新的原则是新的子空间重建信号残差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确保重建,并能确保下一个迭代循环找到更好的子空间。

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/群测试快速获得信号的支撑集,实现精确重构。可将重构问题表述为:已知长度为 N 的矢量 $\mathbf{x}$ 中包含 K 个非零元素($\mathbf{x}_i \neq 0$),但其位置分布未知,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将测试表示为一个二进制矩阵 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为线性关系, 则重构目标稀疏矢量 $\mathbf{x}$ 的问题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, 且不易确定测量矩阵的约束条件, 为构建新的测量矩阵带来了困难。

下面介绍两种简单的组合算法, 均要求测量值无噪声干扰。

(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 $h: \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, m\}$ 的集合, 且为有限集合, 其大小为 m^N 。 H 中每个函数 h 可通过一个大小为 $m \times N$ 的二进制矩阵 $\Phi(h)$ 表示, 其中每一列都是一个二进制矢量, 仅在 $j=h(i)$ 位置为 1。为构建完整的采样矩阵 Φ , 可根据均匀分布从 H 中独立选取 d 个函数 $h_1, h_2, \dots, h_d$, 将其二进制矩阵垂直堆叠为尺寸为 $M \times N$ 的矩阵, 该矩阵即为 Φ , 且每列均有 d 个 1, 其中 $M=md$ 。

针对已知任意信号 $\mathbf{x}$, 可获得线性测量值 $\mathbf{y}=\Phi\mathbf{x}$ 。根据下列两个特性, 很容易得到对测量值的直观认识。首先, 作为测量值矢量的 $\mathbf{y}$ 很容易借助母二进制函数 $h_1, h_2, \dots, h_d$ 得到分组的形式; 其次, 测量值矢量 $\mathbf{y}$ 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, 对于一个固定的信号系数 j , 每个测量值 y_i 都有函数 h 将 x_j 以汇总的形式映射到相同的 i 上。信号重构的目标是从汇总后的观察值 y_i 中, 恢复出原始信号值 x_j 。

当原始信号为正数时, 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 , 可由下式实现:

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

直观地说, 这意味着重建目标信号的第 j 个系数 x'_j 可通过从所有可能有 x_j 介入而形成的测量值中找出幅值最小的观测值(作为 x'_j)得到。该方法的特点是简单、高效。

(2) 计数-中值略图法

计数-中值略图法在目标信号可能为负数时适用。不是取幅值最小的观测值作为 x'_j , 而是取中值。因为对于一个普通信号, 其他 $\mathbf{x}$ 对 $\mathbf{y}$ 的影响可能是正的, 也可能是负的, 而中值最可能是原始信号的值。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 , 可由下式实现:

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

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, 所以对于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, 将其线性组合作为分类函数实现分类。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,这种信号为某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 $\mathbf{x}$ 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, $N(x | m, \sigma^2)$ 表示随机变量 x 服从均值为 m 、方差为 σ^2 的高斯分布。进一步

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

其中, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]$ 。再假设高斯分布中方差的逆(也就是 $\boldsymbol{\alpha}$)也独立并服从参数为 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, $\text{Gamma}(\alpha_i | a, b) = \alpha_i^{a-1} e^{-\alpha_i b} \Gamma(a)^{-1} b^a$,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。通过贝叶斯推理,可得在给定参数 a 和 b 的条件下, $\mathbf{x}$ 的边缘分布是未标准化学生氏分布,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。标准化学生氏分布为

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

随机变量 $r = \mu + st$ 的分布就是未标准化学生氏分布:

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

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

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

则给定测量值 $\mathbf{y}$,即可通过贝叶斯推理结合各种迭代算法,获得 $\mathbf{x}$ 的后验概率分布。

(2) 贝叶斯压缩感知

可从相关向量机(RVM)模型出发考虑贝叶斯压缩感知(BCS)。对于给定观测值 $\mathbf{y}$,可通过将 $\mathbf{y}$ 在给定 $\mathbf{x}$ 的概率分布时对 $\mathbf{x}$ 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)求解各参数。由式(3.4.39)可得

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

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,可得

$$p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2) = \int p(\mathbf{y} | \mathbf{x}, \sigma^2) p(\mathbf{x} | \boldsymbol{\alpha}) d\mathbf{x}$$

$$=(2\pi)^{-N/2} |\sigma^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T|^{-1/2} \exp \left[-\frac{1}{2} \mathbf{y}^T (\sigma^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T)^{-1} \mathbf{y} \right] \quad (3.4.41)$$

其中, $\mathbf{A}$ 为对角矩阵, 其对角线上的元素由 α 组成。

在此过程中, 需要对一个 $N \times N$ 的矩阵求逆, 所以算法复杂度为 $O(N^2)$ 。采用快速边缘似然率最大化的方法可将复杂度降至 $O(NM^2)$ 。它是一种渐进式模型构造法, 可将基函数顺序地加入或删除, 所以能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, 其优点之一是对所估计信号的每个分量提供一个置信区间以帮助了解和判定该估计是否准确。例如, 置信区间太大表明这个估计可能不够可靠。理想情况下希望得到一个估计值且其置信区间较小。

3.4.4 基于深度学习的重构算法

近年来, 提出了许多基于深度学习技术的重构算法。

1. 深度网络化的重构算法

3.4.3 小节介绍了多种典型的重构算法。有些重构算法已被深度网络化, 即用深度网络实现。下面是两种典型的被深度网络化的重构算法。

(1) 交替方向乘法(ADMM)

ADMM 重构算法可用于求解凸优化问题, 通过引入增广拉格朗日函数, 对问题进行分解, 以实现交替优化[Lee 2016]。对 ADMM 重构算法的深度网络化结果称为 ADMM-Net [Yang 2016], 其根据从 ADMM 重构算法的迭代过程中导出的数据流图, 构建包含重构层、卷积层、非线性变换层和乘数更新层的神经网络, 有效地应用于核磁共振图像重构。

(2) 迭代软阈值算法(ISTA)

ISTA 迭代软阈值算法是一种借助梯度下降的思想, 在迭代中通过软阈值操作更新、求解最优解的算法[Beck 2009]。对 ISTA 重构算法的深度网络化结果包括 ISTA-Net、ISTA-Net+、ISTA-Net++。ISTA-Net 将传统压缩感知迭代算法展开为多层神经网络[Zhang 2017a]。ISTA-Net+引入跳跃连接, 图像更加稀疏化, 有助于训练更深层次的网络[Zhang 2018a]。ISTA-Net++更适合需要灵活处理单一模块的多个压缩感知比率的情况[You 2021]。

2. 深度神经网络端到端的学习

还有些重构算法直接利用深度神经网络实现了端到端的学习, 如表 3.4.1 所示[李 2022a]。

表 3.4.1 端到端学习的网络模型

网络模型	典型示例	结构特点	参考文献
基于全连接网络	堆叠去噪自编码器(SDA)	线性测量重构包括 3 层降噪自编码器, 非线性测量重构包括 4 层降噪自编码器	[You 2015]
基于卷积神经网络	ReconNet	使用多层卷积神经网络进行重构, 在降低参数数量的同时增强网络模型的表达能力	[Kulkarni 2016]
基于残差网络	DR ² -Net	由全连接网络生成初始重构图像, 然后通过残差网络进一步提升重构质量并降低时间复杂度	[Yao 2019]
	CSNet	用深度残差卷积神经网络对图像进行采样与重建	[Shi 2019b]
基于生成对抗网络	DAGAN	生成器基于 U-Net, 包括 8 个卷积层和 8 个去卷积层; 鉴别器包括 11 个卷积层	[Yang 2018]

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基本原子的加权组合构成,对应信号的稀疏表达。字典学习的目标是估计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。

可采用数学模型描述字典学习。为简化计算仅考虑实数信号。设 $\mathbf{y}_i \in \mathbb{R}^L$ 是一个维度为 L 的矢量,用于表示 N 个信号中的第 i 个; $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_M] \in \mathbb{R}^{L \times M}$ 表示原子信号矩阵,其中第 i 列是维度为 L 的原子信号; $\mathbf{w}_i \in \mathbb{R}^M$ 为维度为 M 的矢量,用于表示构成矢量 $\mathbf{y}_i$ 的原子信号的权重矢量。根据前面的假设,有

$$\mathbf{y}_i = \mathbf{A}\mathbf{w}_i + \boldsymbol{\epsilon}_i \quad \|\mathbf{w}_i\|_0 \leq K \quad i = 1, 2, \dots, N \quad (3.5.1)$$

其中, $\boldsymbol{\epsilon}_i \in \mathbb{R}^L$ 是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 是大于零的正整数。

式(3.5.1)表达了多个重要信息。

(1) 信号是原子矢量的线性组合,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,缺点是不易描述非线性数据。不过这里并没有对信号所在的空间进行限制,所以如果要引入非线性成分,可先将信号转换到某个特征空间,再使用线性模型。

(2) 非确定性由误差项 $\boldsymbol{\epsilon}_i$ 体现。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,即可得到 PCA(参见中册 A.6.2 小节)模型。

(3) 原子信号的权重矢量 $\mathbf{w}_i$ 是 K 稀疏的,这由 $\|\mathbf{w}_i\|_0 \leq K$ 表达,表明感兴趣的信号是由原子信号构成的。需要指出,稀疏性假设在字典学习中并非不可或缺。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。

可将式(3.5.1)表达为矩阵形式:

$$\mathbf{Y} = \mathbf{A}\mathbf{W} + \mathbf{E} \quad \|\mathbf{w}_i\|_0 \leq K \quad i = 1, 2, \dots, N \quad (3.5.2)$$

其中, $\mathbf{Y} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \dots \quad \mathbf{y}_N]$, $\mathbf{W} = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \dots \quad \mathbf{w}_N]$, $\mathbf{E} = [\boldsymbol{\epsilon}_1 \quad \boldsymbol{\epsilon}_2 \quad \dots \quad \boldsymbol{\epsilon}_N]$ 。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稀疏,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,且待测信号只有一个样本。但在字典学习中,仅已知多个观测信号,而未知参数却有两个,即原子信号矩阵 $\mathbf{A}$ 及权重矩阵 $\mathbf{W}$,并需要考虑多个样本。可见,字典学习更困难,求解也更复杂。

3.5.1 字典学习与矩阵分解

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)可简化为

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

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

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

其中, $\mathbf{U} \in \mathbb{R}^{L \times L}$ 是一个包含正交单位列矢量的矩阵, 即 $(i, j = 1, 2, \dots, L)$

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

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

$\mathbf{D} \in \mathbb{R}^{L \times N}$ 是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, $\mathbf{D}$ 中第 i 行第 j 列元素

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

d_{ii} 为矩阵 $\mathbf{Y}$ 的第 i 个奇异值, 且如果 $i > j$, 则有 $d_{ii} \geq d_{jj}$, 即主对角线上的奇异值按降序排列。

$\mathbf{V} \in \mathbb{R}^{N \times N}$ 与 $\mathbf{U}$ 一样, 是一个包含正交单位列矢量的矩阵。

将矩阵 $\mathbf{Y}$ 的秩记为 $R(\mathbf{Y})$, 则 $R(\mathbf{Y}) \leq \min\{L, N\}$ 。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 $R(\mathbf{Y}) = T$, 可以截掉 SVD 分解中各矩阵冗余的列, 得到较简洁的 SVD 分解矩阵, 即 $\mathbf{U} \in \mathbb{R}^{L \times T}$, $\mathbf{D} \in \mathbb{R}^{T \times T}$, 且 $\mathbf{V} \in \mathbb{R}^{N \times T}$, 这通常称为瘦 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) 字典学习问题的一个解。但实际上式 (3.5.3) 中为约等号, 说明存在误差。而式 (3.5.4) 的 SVD 解没有考虑误差, 所以不符合实际情况。解决的方法是保留若干重要的奇异值, 将不重要的奇异值设为零。从而得到截断 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 是矩阵 $\mathbf{Y}$ 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 保留几个最大的奇异值 (设保留了 T 个), 舍弃其余的奇异值以逼近矩阵 $\mathbf{Y}$:

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

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

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

所以采用 tSVD 方法的字典学习方案可显式地表示误差, $\mathbf{A} = [\mathbf{u}_1 \cdots \mathbf{u}_N]$, $\mathbf{W} = [d_1 \mathbf{v}_1 \cdots d_T \mathbf{v}_T]$ 。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

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

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

残差在原始信号矩阵 $\mathbf{Y}$ 中所占的比例为

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

可见, 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)最大奇异值重建信号。

为确定 T , 可根据 SVD 与 PCA 之间的关系, 即 SVD 的 $\mathbf{U}$ 矩阵对应 PCA 中的主分量, $\mathbf{D}\mathbf{V}^T$ 就是数据在主分量上的投影值。可看出, 式(3.5.8)的 tSVD 逼近实际上要选取若干主分量重构信号。所以, 可以利用 PCA 中确定主分量个数的方法确定 tSVD 中的 T (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表示(第一行是待优化目标, 其下两行是限定条件):

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

其中, $\mathbf{I}_M$ 是 $M \times M$ 的单位矩阵。

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

式(3.5.13)存在两个问题。首先, tSVD 是其一个解, 但由于没有限制 $\mathbf{W}$ 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 $\mathbf{A}^*$ 和 $\mathbf{W}^*$ 是一组解, $\mathbf{H}$ 是一个 $M \times M$ 的西矩阵($\mathbf{H}^T \mathbf{H} = \mathbf{H}\mathbf{H}^T = \mathbf{I}_M$), 则 $\mathbf{A}^* \mathbf{H}$ 和 $\mathbf{W}^* \mathbf{H}$ 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。

由式(3.5.13)可见, 如果要去掉原子信号必须正交的限制, 只要去掉第 3 行即可。但这可能导致解的空间太大及平凡解的产生。所以需要限制 $\mathbf{A}$ 或 $\mathbf{W}$ 或两者同时进行适当限制。用于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, 可满足上述两个条件。实际信号具有物理意义, 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 的限定条件是原子信号和权重矩阵必须都为非负。NMF 的优化形式类似于式(3.5.13), 可写为

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

考虑到优化目标的多样性, 还可写为更具普适性的形式:

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

其中, $d(\mathbf{P}, \mathbf{Q})$ 是表示矩阵 $\mathbf{P}$ 和 $\mathbf{Q}$ 非相似度的函数。设

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

则式(3.5.15)可化为式(3.5.14)的形式。

NMF 的解空间一般偏大,常常增加一些限制条件以进一步缩小解空间。例如,可加入权重矩阵的稀疏性条件以实现稀疏非负矩阵分解:

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

另外,对流形上的非负矩阵分解,可以有

$$\begin{aligned} \min_{\mathbf{A} \in \mathbf{R}^{D \times M}, \mathbf{W} \in \mathbf{R}^{M \times N}} & d(\mathbf{Y}, \mathbf{AW}) \\ \text{s. t. } & \mathbf{Y} = \mathbf{AW} + \mathbf{E} \\ & \mathbf{A} \geq 0 \\ & \mathbf{W} \geq 0 \\ & \mathbf{a}_i \in S \quad i = 1, 2, \dots, S \end{aligned} \quad (3.5.18)$$

其中, S 表示一个具有特定结构的流形。

3.5.3 端元提取

端元提取也是一种字典学习方法,其两个限制条件为:①权重矩阵非负;②权重矩阵每一列的和必须为1,即

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

其中, $\mathbf{L}$ 为一个长度合适的全为1的列矢量。将式(3.5.19)写为矩阵形式:

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

一种典型的端元提取优化方法是迭代约束端元化(ICE):

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

其中, $\mathbf{A}$ 为端元矩阵,其第 i 列 $\mathbf{a}_i$ 为第 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,用于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,但优化过程不同。ICE 采用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 中,固定一个未知量,如 $\mathbf{W}^t$,求解另一个未知量,如 $\mathbf{A}^{t+1}$;得到 $\mathbf{A}^{t+1}$ 后再固定 $\mathbf{A}^{t+1}$,求解 $\mathbf{W}^{t+1}$ 。循环迭代即可得到最终解。

具体将 ICE 的优化问题简化为

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

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

$$\sum_{i \neq j} \|a_i - a_j\|_2^2 = \text{tr}(\mathbf{A}\mathbf{H}\mathbf{A}^T) \quad (3.5.23)$$

另外定义

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

如果第 t 步迭代已得到 $\mathbf{A}^t$ 和 $\mathbf{W}^t$, 则 $\mathbf{W}^{t+1}$ 可由(固定 $\mathbf{A}$)下式求解得到:

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

式(3.5.25)是一个带线性限制条件的二次规划(QP)问题,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 $\mathbf{A}$:

$$\mathbf{A}^{t+1} = \arg\min_{\mathbf{A}} J(\mathbf{A}, \mathbf{W}^{t+1}) \quad (3.5.26)$$

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 $M-1$ 。只有在某些情况下(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 的优化目标同样是凸性的,所以上述算法虽然可以保证收敛,但可能只收敛到局部最优解,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)],也可对式(3.5.21)进行相应的变形,以克服 ICE 的缺点或引入 ICE 中没有的性质。例如,可得到稀疏迭代约束端元化(SPICE):

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

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

稀疏编码是寻找数据的一种字典表达方式,要使每个数据都可表示为少数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为

$$\begin{aligned} \min_{\mathbf{A} \in \mathbf{R}^{L \times M}, \mathbf{W} \in \mathbf{R}^{M \times N}} & \|\mathbf{E}\|_p \\ \text{s. t. } \mathbf{Y} &= \mathbf{A}\mathbf{W} + \mathbf{E} \\ \|w_i\|_0 &\leq K \quad i=1, 2, \dots, N \end{aligned} \quad (3.5.29)$$

式(3.5.29)表达的含义是找到用于表示数据 $\mathbf{Y}$ 的字典,确保使用该字典的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相似之处,但有两点不同:一是压缩感知中的讨论针对信号的感知与重建;二是压缩感知中的采样矩阵 Φ 对应稀疏编码中的矩阵 $\mathbf{A}$,且 Φ 已知,但在稀疏编码中矩阵 $\mathbf{A}$ 也是需要求解的目标之一。

可见,稀疏编码对优化问题求解的复杂度远超压缩感知。

1. 最优方向法

最优方向法(MOD)解决的是式(3.5.29)的问题并实现以下稍许不同的目标:

$$\begin{aligned} \min_{\mathbf{A} \in \mathbb{R}^{L \times M}, \mathbf{W} \in \mathbb{R}^{M \times N}} \|\mathbf{W}\|_0 \\ \text{s. t. } \mathbf{Y} = \mathbf{A}\mathbf{W} + \mathbf{E} \\ \|\mathbf{e}_i\|_2 \leq \epsilon \quad i = 1, 2, \dots, N \end{aligned} \quad (3.5.30)$$

其中, $\|\mathbf{W}\|_0$ 是矩阵 $\mathbf{W}$ 的 L_0 范数,即矩阵 $\mathbf{W}$ 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,主要包括两个步骤。

(1) 稀疏编码阶段: 对于 $i = 1, 2, \dots, N$, 使用匹配追踪算法以获取

$$\mathbf{w}_i = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{y}_i - \mathbf{A}^{\prime-1} \mathbf{w}\|_2^2 \quad \text{s. t. } \|\mathbf{w}\|_0 \leq K \quad (3.5.31)$$

(2) 字典更新阶段: 使用下列公式更新原子信号矩阵:

$$\mathbf{A}^t = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{A}\mathbf{W}^t\|_F^2 = \mathbf{Y}\mathbf{W}^t [\mathbf{W}^t (\mathbf{W}^t)^T]^{-1} \quad (3.5.32)$$

2. 核奇异值分解

在 MOD 中,字典更新通过对权重矩阵进行简单的伪逆计算实现,使一些原子信号混入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)为消除原子信号之间的互相干扰而提出。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 $\mathbf{a}_j$, 令 S_j 为包含原子信号 $\mathbf{a}_j$ 中数据的索引集合,即

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

w_{ji} 为权重矩阵 $\mathbf{W}$ 的第 j 行第 i 列元素。因为 $\mathbf{W}$ 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,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:

$$\mathbf{E}_j = \mathbf{Y} - \sum_{l \neq j} \mathbf{a}_l \mathbf{w}_l^T \quad (3.5.34)$$

上式计算的是除 $\mathbf{a}_j$ 外其他原子信号的拟合残差。引入记号 $\mathbf{E}_j^S$ 以表示矩阵 $\mathbf{E}_j$ 的列子矩阵,即由 $\mathbf{E}_j$ 中属于 S 的列组成的矩阵。对 $\mathbf{a}_j$ 和 $\mathbf{w}_j$ 的更新由下式求解得到:

$$(\mathbf{a}_j, \mathbf{w}_j) = \underset{\mathbf{a}, \mathbf{w}}{\operatorname{argmin}} \|\mathbf{E}_j^S - \mathbf{a}\mathbf{w}^T\|_F^2 \quad (3.5.35)$$

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 $\mathbf{a}$ 和 $\mathbf{w}$ 都是矢量,而非矩阵。所以,可用经典的矩阵分解求解。在 K-SVD 中使用对矩阵 $\mathbf{E}_j^S$ 的秩为 1 的 SVD 分解,即 tSVD

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

可见, $\mathbf{u}$ 和 $\lambda\mathbf{v}$ 分别为 $\mathbf{a}$ 和 $\mathbf{w}$ 的解。如果 $K=1$ 且限定权重矩阵的值只能取 1 或 0,那么原始问题简化为标准的 K 均值聚类问题。类似于在每个 K 均值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寻找 K 个子矩阵(对应 K 个子集)。其两个主要迭代步骤如下。

(1) 稀疏编码阶段: 对于 $i = 1, 2, \dots, N$, 使用匹配追踪算法以获取

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

的近似解,得到 $\mathbf{W}^t$ 。

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

$1, 2, \dots, M)$ 。

先计算残差矩阵:

$$E_j = Y - \sum_{l \neq j} a_l w_l^T \quad (3.5.38)$$

而矩阵 E_j^S 的秩为 1 的 tSVD 分解为

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

其中, $a_j = u$, 且权重矢量 $w_j = \lambda v$ 。

可见, 稀疏编码与聚类分析较为相似, 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

压缩感知已得到许多实际应用, 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一个使压缩感知得到广泛关注的重要事件。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。其成像流程如图 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。这是一种借助大量微小镜片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正负 12° 的机械翻转, 对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, 低光照时也可使用血崩二极管)接收, 其输出电压随反射光强度变化。对该电压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作一个矢量 x , 则该次测量的结果 $y_i = \phi_i x$ 。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, 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数的 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过程中实现对图像数据的压缩。

例 3.6.1 单像素相机成像效果

目前单像素相机成像的效果与普通 CCD 相机成像的效果还有一定的距离。图 3.6.2 给出了一组示例。图 3.6.2(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 3.6.2(b)是用单像素相机成像的一种效果, 此时 M 为像素个数的 16%(进行了 11 000 次测量); 图 3.6.2(c)是用单像素相机成像的另一种效果, 此时 M 为像素个数的 2%(进行了 1300 次测量)。

由图 3.6.2 可见, 虽然测量数量有所减少, 但质量还不能比较。另外, 机械翻转微镜需要一定的时间, 所以获得足够多测量数量的成像时间较长(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, 单像素相机比较有优势。由于红外谱段的探测器件较昂贵, 单像素相机可与其竞争。

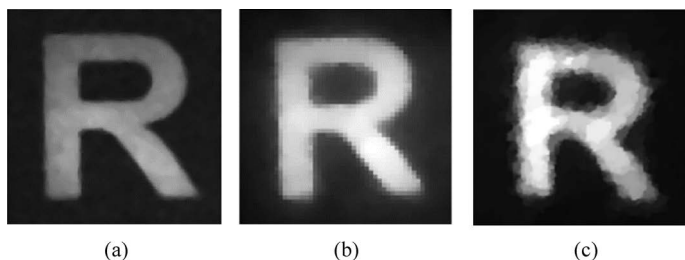


图 3.6.2 单像素相机成像效果示例

□

单像素相机成像的基础是计算鬼成像理论[Shapiro 2008]。根据计算鬼成像理论,在鬼成像过程中,自由传播的参考光束的测量信号无须来自真实物理光束,且无须探测器探测,可通过计算获得;而景物的透射光束信号用一个桶探测器(没有空间分辨率)接收;两路信号进行相关运算即可获得景物透光率的空间分布,即获得图像。如果采用多个不同位置的单像素相机同时成像,并借助从影调恢复形状的技术(参见第9章),就可实现单像素3-D成像[Sun 2013]。进一步,[Edgar 2015]提出了单像素动态成像技术,并获得了可见光和红外视频图像。通过考虑环境电磁辐射因素,依据景物信息在空域与频域的傅里叶变换关系,[陈 2021b]推导了单色单像素频谱重构成像理论,提出了实现单像素彩色成像的3种方案。

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获取图像的方法(参见上册第9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图像采集需要大量时间,不仅增加了患者的不适感,而且患者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影响。因此,缩短成像时间、提高成像效率和减少图像伪影受到广泛关注。

利用压缩感知的磁共振成像的基本思路如下:先对图像采用离散傅里叶标准正交基进行稀疏表达,再对得到的 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图像。整个流程如图 3.6.3 所示。

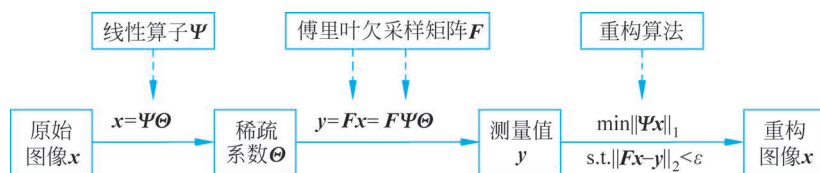


图 3.6.3 压缩感知磁共振成像流程

具体描述如下。

(1) 考虑一幅 N -D 图像 x 在 $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_N\}$ 下具有稀疏性,则

$$x = \sum_{i=1}^K \theta_i \Psi_i = \Psi \Theta \quad (3.6.1)$$

其中, $\Theta = [\theta_1 \ \theta_2 \ \dots \ \theta_N]^T$ ($K \ll N$) 代表 x 在稀疏变换 Ψ 中的非零系数。

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

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

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精确重构:

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

式(3.6.3)表示在用不等式约束 K 空间数据一致性条件下(ϵ 代表测量值重构的保真度)通过稀疏变换 Ψ 对 $\mathbf{x}$ 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 给出了压缩感知磁共振成像效果示例。图像为用纵向松弛时间 T_1 加权的脑图像,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,测量次数分别为原始像素数的 $1/4$ 、 $1/6$ 和 $1/8$ (都对应欠采样)。换句话说,成像的加速倍率(反比于测量次数)分别为 4、6 和 8。

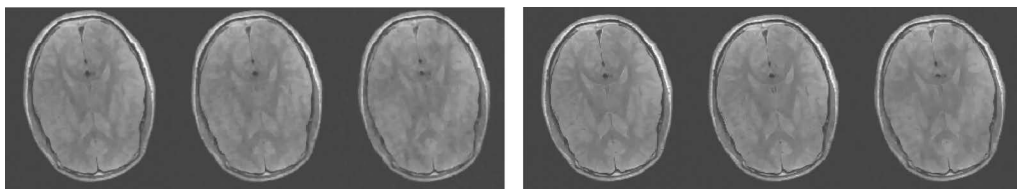


图 3.6.4 压缩感知磁共振成像效果示例

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缺失。然后,通过仅利用 K 空间的部分数据就可成功重构出原始图像。如此,不但能加快成像速度,而且能极大减少扫描时间并提高患者的舒适度。

总结和复习

随堂测试

