DOI:10.13203/j.whugis20130267
文章编号:1671-8860(2014)11-1261-08

压缩感知的信息论解译

张景雄[1] 阳 柯[1] 郭建中[2]

1 武汉大学遥感信息工程学院,湖北 武汉,430079
2 武汉纺织大学电子与电气工程学院,湖北 武汉,430074

摘 要:压缩感知(compressed sensing or compressive sampling, CS)是数据采集与信号重构的新体制,其与信息论的关系是,应该且可以从信息论的角度对CS进行分析,而CS丰富和发展着信息论的内涵和外延。换言之,信息论的一些基本概念和原理(如信源、信道、信源编码、信道编码、率失真、Fano不等式、数据处理定理等)为CS研究提供了理论基础,尤其是在性能限(如采样数)的界定等方面;另一方面,CS提供了采集、存贮、传输、恢复稀疏信号的高效方法,以其独特的理念和算法模式,提供了直接对信息的采样和处理机制,延拓了经典信息论的范畴。本文将梳理和阐释CS和信息论之间的关系,力图从信息论角度揭示CS中的一些基本问题,尤其是CS采样问题,并寻求用信息论指导CS的学习与研究。

关键词:压缩感知;信息论;稀疏;采样;熵

中图法分类号:P208 **文献标志码**:A

信息论[1]是运用概率论与数理统计的方法研究信息、信息熵、通信系统、数据传输、数据压缩等问题的应用数学学科,涵盖了信息度量、编码理论、算法复杂性理论和算法信息论等方向。信息传输和信息压缩是信息论研究中的两大领域,这两个方面又由信道编码定理、信源-信道隔离定理相互联系[2]。信息论已广泛应用于许多学科领域,包括测绘、遥感(光学与微波)和地理信息,提供了地理空间信息的理论支撑[3-4]。

数据采集也称信号采样,是信息论的重要研究内容,它是模拟信号至数码数据转换的必由之路。Shannon-Nyquist采样定理要求,对一带宽为B的连续带限信号,需以至少$2B$的采样率(即$1/(2B)$的采样间距)来采样获取离散化的信号,以使原连续信号能从等间隔离散化的信号通过sinc函数加权完全恢复。由于分辨率的持续提高和数据量的剧增,为了降低存储、处理和传输的成本,需要对信源信号进行压缩处理。传统压缩编码往往通过变换来实现压缩,但这种采样后再压缩信号的方法效率很低。

信息社会、大数据时代的数据和信息需求与日俱增,加剧了数据采集、传输和存储的压力,使得数据采集的效率与精度同等重要,传统采样-压缩的模式弊端凸显。Nyquist采样率是原连续信号能完全恢复的充分条件,但并非必要条件,可以运用信号的其他特性,以提高采样效率。压缩感知(CS)理论指出,若信号是稀疏的(在某个变换域)或可压缩的,可通过不相关的观测矩阵将其(高维信号)投影到低维空间,进而利用该投影(低维观测)求解优化问题,以高概率地重构原信号[5-8]。压缩感知和变换编码都实现了对信源的压缩,不同之处在于变换编码先用Nyquist速率进行采样,再变换去相关性,而压缩感知将去相关性和采样融为一体;Nyquist采样后信号的重构是通过简单的线性加权平均实现的,而压缩感知的重构基于优化算法(非线性的)。

CS是信号采集的新理论,前提是信号的稀疏性。CS信号采集过程从Analog-Data模式变成了Analog-Information模式,于是CS可以理解为直接感知压缩了的信息。若想只采集少部分数据并从中解压缩出大量信息,需保证这些以sub-Nyquist采样率得到的少量的数据包含了原信号的全局信息及存在某种算法能从这些数据中还原出原有的信息。在某些场合,上述第一要求是自动满足的,如医学图像成像如核磁共振(MRI)技术,仪器所采集到的不是直接的图像像素,而是图

收稿日期:2013-06-24

项目来源:国家973计划资助项目(2010CB731905);国家自然科学基金资助项目(41071286);湖北省教育厅基金资助项目(D201416022)。

第一作者:张景雄,博士,教授。主要从事遥感与地理信息的基础研究和应用研究。E-mail: jxzhang@whu.edu.cn

像经傅立叶变换后的数据,每一数据都在某种程度上包含了全图像的信息,每部分图像信息"弥漫"于每单个数据里。上述第二点的理论依据是Candes等学者的工作:如果假定信号满足某种特定的稀疏性,可以从少量的测量数据中,高概率的还原出原始的信号来,其中所需要的计算部分是一个复杂的迭代优化过程,如所谓的ι_1-最小化算法。

CS的原理简述如下。设信号$\boldsymbol{X}=(X_1, X_2, \cdots, X_N)^{\mathrm{T}}$是稀疏的,则它可以用基$\Psi$($N\times N$维矩阵)表示为:

$$\boldsymbol{X} = \boldsymbol{\Psi\Theta} \tag{1}$$

显然向量$\boldsymbol{X}$和向量$\boldsymbol{\Theta}$($\Theta=(\theta_1,\theta_2,\cdots,\theta_N)^T$)是同一个信号的等价表示。如果向量$\Theta$中仅仅有$k$个非零分量,且$k\leqslant N$,则称信号$\boldsymbol{X}$或$\boldsymbol{\Theta}$为$k$-稀疏。

设矩阵$\boldsymbol{A}$($\boldsymbol{M}\times \boldsymbol{N}$维矩阵)和基$\Psi$不相关,则无噪声干扰的压缩感知的观测模型可表示为:

$$\boldsymbol{Y} = \boldsymbol{AX} \tag{2}$$

式中,$\boldsymbol{X}$是未知的N维向量;$\boldsymbol{Y}$是M维向量。借助M个观测值$\boldsymbol{Y}=(Y_1, Y_2, \cdots, Y_M)^{\mathrm{T}}$和$k$-稀疏度的约束条件,通过凸优化或线性规划重构信号$\boldsymbol{X}$[9]。

与Shannon信息论相似,CS理论与应用的一个重要课题是采样数的确立[10-13]。本文将以Fourier系数为观测矩阵的情形为例,比较Nyquist采样与CS采样的联系与区别,以深层次理解本文的主旨。然后,文章讨论观测矩阵的信息探测效率,并分别就CS与信源编码、信道编码、率失真、Fano不等式等概念和原理的联系展开讨论。最后,对相关理论和应用问题进行概括。

1 Shannon-Nyquist数据采样和CS信息采样

Shannon-Nyquist采样定理给出了带限信号的重构的充分条件,即至少两倍带宽(设带宽为B_0)的Nyquist采样率($2B_0$)采集信号,从而可以利用sinc函数精确恢复原信号$\boldsymbol{X}(t)$:$\boldsymbol{X}(t)=\sum_{i=-\infty}^{\infty}\boldsymbol{X}(i/2B_0)\cdot\frac{\sin[2\pi B_0(t-i/2B_0)]}{2\pi B_0(t-i/2B_0)}$。带宽是一关键概念:在无线电通信中,带宽是调制载波占据的频率范围;在数学中,对于可以看作时间函数的模拟信号来说,带宽是以赫兹为单位、信号的傅里叶变换不为0的频率范围(这对应于数学里函数的支撑的术语,即函数不为零的(一维)定义域的长度)。在离散时间系统和信号处理中,采样率与带宽有关;根据Shannon-Hartley定理,可靠通信的数据速率与通信所用信号频率范围成比例,不论该频带在频谱的位置。

CS环境下,采样率的界定是基于稀疏度和观测矩阵特性等参数的。如下文所述,经典的Nyquist采样定理是基于Fourier变换和频谱的有限特性的,而此环境下的CS采样的必要观测数与信号的稀疏度成正比。

考虑一个N长的离散信号$\boldsymbol{X}$(即N个元素的向量):$\boldsymbol{X}$在频率域的支撑集为$\boldsymbol{\Omega}$,该支撑集的测度(即$\boldsymbol{\Omega}$元素个数)为B,即$B=|\boldsymbol{\Omega}|$。若$\boldsymbol{\Omega}$是连通的,可以将B看作是$\boldsymbol{X}$的带宽(即认为$B=k$,前已述及的稀疏非零元素的个数);当$\boldsymbol{\Omega}$已知时,Shannon采样定理可以适用,即$\boldsymbol{X}$可以完全重构的条件是采集B个等间隔的样本,重构方法是简单的sinc核的线性内插。当$\boldsymbol{\Omega}$未知(虽然该集合的测度仍为B,对于稀疏信号,B为其非零元素个数),X完全恢复的至少采样数(Fourier变换域)为$B\log N$(这些样本的选择是任意的,只要样本数符合要求,其中的对数的底缺省为2),且$\boldsymbol{X}$的恢复转化为P_1问题的求解[14]。从信息论的角度看,每个稀疏信号的编码需$\log N$bits,所以k-稀疏的信号的编码需$k\log N(=B\log N)$bits。

设所获取的观测数据$\boldsymbol{Y}=\{y_m\}$($m=1,\cdots,M$)是信号$\boldsymbol{X}$的离散Fourier变换系数:

$$y_m = \sum_{i=1}^{N} x_i \exp(-\mathrm{i}2\pi(i-1)m/N)$$

式中,$m\in\{1,2,\cdots,M\}$;$\mathrm{i}^2=-1$。其向量和矩阵的形式为:

$$\boldsymbol{Y} = \boldsymbol{FX} \tag{3}$$

式中,F的元素是离散Fourier变换系数,即$\boldsymbol{F}(m,n)=\{\exp(-\mathrm{i}2\pi(n-1)m/N\}$。从$\boldsymbol{Y}$恢复稀疏信号$\boldsymbol{X}$其实是一个凸优化过程:

$$\min_{\boldsymbol{X}} \| X \|_{l_1} \quad \mathrm{s.\,t.} \quad \boldsymbol{FX} = \boldsymbol{Y} \tag{4}$$

可以看出,类似于Shannon采样定理,稀疏信号的采样数也可以通过离散Fourier变换来确定;不同于Shannon采样定理,决定CS采样数的参数实质是稀疏度,虽然稀疏度使用了与带宽同样的变量$\boldsymbol{B}$。

Nyquist采样与CS采样还可以延伸分析。根据Peleg与Shamai[15]的论述,可以将采样情形与相关使用的结论归类。设$\boldsymbol{X}(t)$是一带宽为B的连续信号,其采样后的离散信号为$\boldsymbol{Y}(m)=\boldsymbol{X}(t_m)$。采样的时间(空间)坐标$t_m$可能是不规则

的，平均采样率为 R。信号重构所需的采样率取决于 $\boldsymbol{X}$ 的功率谱密度 PSD（以及知否该 PSD）、信噪比 SNR、所需的重构可靠性和重构算法的复杂度。若 $\boldsymbol{Y}$ 的频谱支撑是非连通、多通带的，它们的带宽分别为 B_1，B_2，…，B_b（b 为通带数）。可以简单计算等效带宽：

$$B^* = \sum_{j=1}^{b} B_j$$

以及采样率与等效带宽的比值：

$$r = R/B^* \tag{5}$$

b 与 r 的取值不同，对应的采样方式也不同。当 $b=1$，$r=1$ 时，为 Shannon-Nyquist 采样；$r=1$ 对应 Landau 采样；当 $r=2+\varepsilon'$（$\varepsilon'>0$）时为 CS 采样。

Fourier 变换系数组成的矩阵只是 CS 观测矩阵取值的一种特例。下文将讨论 CS 观测矩阵的若干一般特性，因为它们有助于确定必要的采样数。

2　观测矩阵的信息探测效率

CS 的一个前提是信号 $\boldsymbol{X}$ 的稀疏性（信号越稀疏，所需要的观测值的个数 M 就越少）；另一个前提是观测矩阵 $\boldsymbol{A}$ 的信息感知能力。

CS 环境下，观测值是信号 $\boldsymbol{X}$ 在一组投影向量上的线性投影；投影向量的作用是一组探测包含于 $\boldsymbol{X}$ 的信息的探测器[16]。投影向量即对应观测矩阵的列。随机观测矩阵具有信息探测的优良品质，成为稀疏信号采集的最优策略，因为它只需近乎最少的观测数[9]。

在无噪声条件下，压缩感知模型可以表示为式(2)。可以设 $\boldsymbol{X}$ 在某个域（稀疏基 Ψ）是稀疏的；有时为了方便起见，可设 $\boldsymbol{X}$ 本身就是稀疏的。以下叙述无噪模型下的采样数条件，包括 $\boldsymbol{A}$ 的 Spark 和相关性。

观测矩阵 $\boldsymbol{A}$ 的一定 Spark 数量可以确保 $\boldsymbol{Y}$ 至多与满足 $\boldsymbol{Y}=\boldsymbol{AX}$ 的一个向量匹配。对于观测矩阵 $\boldsymbol{A}$，Spark($\boldsymbol{A}$)定义为矩阵 $\boldsymbol{A}$ 中列向量组线性相关的最小的列数。Spark($\boldsymbol{A}$)的取值范围为[2，$M+1$]。研究表明[16]，对于任意向量 $\boldsymbol{Y}=\boldsymbol{AX}$，可以精确恢复 $\boldsymbol{X}$ 的充要条件（即最多存在一个 k-稀疏信号 $\boldsymbol{X}$，使得 $\boldsymbol{Y}=\boldsymbol{AX}$）是：

$$\mathrm{Spark}(\boldsymbol{A}) > 2k \tag{6a}$$

这样，可以推断出得到最少采样数（即矩阵 $\boldsymbol{A}$ 的行数）：

$$M \geqslant 2k \tag{6b}$$

Spark 的计算极其繁复，自相关特性分析相对容易。观测矩阵 $\boldsymbol{A}$ 的自相关 $\mu(\boldsymbol{A})$ 定义为：

$$\mu(\boldsymbol{A}) = \max_{1\leqslant i\neq j\leqslant M} \frac{|\langle \boldsymbol{A}_i, \boldsymbol{A}_j\rangle|}{\|\boldsymbol{A}_i\|_2 \|\boldsymbol{A}_j\|_2} \tag{7}$$

式中，运算符 $\langle \boldsymbol{A}_i, \boldsymbol{A}_j\rangle$ 表示矩阵 $\boldsymbol{A}$ 的第 i，j 行向量的内积；$\|\cdot\|_2$ 表示某向量的 L_2 范式。$\mu(\boldsymbol{A})$ 取值范围为[$\sqrt{(N-M)/M(N-1)}$，1]，其与 Spark 的关系如下：

$$\mathrm{Spark}(\boldsymbol{A}) \geqslant 1 + \frac{1}{\mu(\boldsymbol{A})} \tag{8a}$$

结合条件式(6a)，有：

$$k < \frac{1}{2}\left(1+\frac{1}{\mu(\boldsymbol{A})}\right) \tag{8b}$$

研究表明，在满足上述条件时，最多只有一个 k-稀疏向量 $\boldsymbol{X}$ 满足 $\boldsymbol{Y}=\boldsymbol{AX}$[16]。

对于本身不稀疏，而在稀疏基 $\boldsymbol{\Psi}$ 上稀疏的信号 $\boldsymbol{X}$，可以定义观测矩阵 $\boldsymbol{A}$ 和稀疏基 $\boldsymbol{\Psi}$ 之间的互相关 $\mu(\boldsymbol{A}, \boldsymbol{\Psi})$：

$$\mu(\boldsymbol{A}, \boldsymbol{\Psi}) = \max_{1\leqslant i\leqslant M, 1\leqslant j\leqslant N} |\langle \boldsymbol{A}_i, \boldsymbol{\Psi}_j\rangle| \tag{9}$$

并给出了最小观测值个数[14]：

$$M \geqslant ckN\mu^2(\boldsymbol{A}, \boldsymbol{\Psi})\log(N/\delta) \tag{10}$$

式中，c 为常数，δ 为反应误差大小的另一常数。在 M 满足上述表达式时，重构算法能以 $1-\delta$ 的概率重构原信号[6,16]，常数 $\delta<1$。

$\mu(\boldsymbol{A})$ 可理解为 $\boldsymbol{A}$ 的行（或列）向量的浓缩程度的度量；每行（或列）向量的 L_2 范数为 $\sqrt{N}$，$\mu(\boldsymbol{A})$ 的值域范围为[1，$\sqrt{N}$]，前者对应一个完全水平的 $\boldsymbol{A}$ 的元素，后者对应每行的"重量"集中于一个元素。对于在 $\boldsymbol{\Psi}$ 上稀疏的信号，观测矩阵 $\boldsymbol{A}$ 和稀疏基 $\boldsymbol{\Psi}$ 之间的互相关 $\mu(\boldsymbol{A}, \boldsymbol{\Psi})$ 粗略地量化了稀疏度与观测体制间的相似性（最小值 1 对应于 $\boldsymbol{A}$ 的每一行向量在 $\boldsymbol{\Psi}$ 域里均匀分布），故而也称为感测模式[9]。

然而，在有噪情形下，需讨论 RIP（restricted isometry property）特性。这是指 $\boldsymbol{Y}=\boldsymbol{AX}$ 与 $\boldsymbol{Y}'=\boldsymbol{AX}'$ 的距离与原信号 $\boldsymbol{X}$ 和 $\boldsymbol{X}'$ 间的距离成正比。给定观测矩阵 $\boldsymbol{A}$ 和一个常数 $\delta\in(0,1)$，对于任一 k-稀疏 $\boldsymbol{X}$，如果满足：

$$(1-\delta_k)\|\boldsymbol{X}\|_2^2 \leqslant \|\boldsymbol{AX}\|_2^2 \leqslant (1+\delta_k)\|\boldsymbol{X}\|_2^2 \tag{11}$$

则称矩阵 $\boldsymbol{A}$ 满足(k，δ)-RIP 性质。RIP 性质可以解释为 $\boldsymbol{A}$ 在编码过程中使得任意一对 k-稀疏向量之间保持了一定的距离，以至于在噪声干扰下，任何两个 k-稀疏向量不会投影为同一个向量。RIP 确保所有 $\boldsymbol{A}$ 的维数为 $M\times k$ 的子矩阵近似保距[16]。可从信息论的角度对 RIP 进行解释。

在有噪声的情况下，编码之后的码字（调制后），被噪声干扰，只要码字（向量）之间保持一定的距离，干扰出错后还能正确译码。

RIP 与相关性的联系通过 Gershgorin 圆定理而获得：若观测矩阵 $\boldsymbol{A}$（列归一化）的相关性为 $\mu(\boldsymbol{A})$，则 $\boldsymbol{A}$ 的 k 阶 RIP 参数 $\delta\leqslant(k-1)\mu$。对于一个 k-稀疏信号 $\boldsymbol{X}$，其恢复的充要条件是 $\boldsymbol{A}$ 满足 $(2k,\delta)$-RIP 条件，因为这意味着 $\mathrm{Spark}(\boldsymbol{A})>2k$。

可以证明，对于元素独立、同分布（如高斯、贝努力）的随机观测矩阵 $\boldsymbol{A}$，其高概率地满足 (k,δ)-RIP 的条件[16]是：

$$M=O(k\log(N/k)/\delta^2) \tag{12}$$

3 CS 与信息论的联系

本文将从信息论的角度讨论信号的稀疏表示、观测、重构（失真度、算法复杂度）等问题，重点是 CS 与信源编码、信道编码、率失真、Fano 不等式的关系。

3.1 CS 与信源及信源编码

对于某一随机变量 $\boldsymbol{X}$，其 Shannon 信息熵（或自信息）的定义是：

$$H(\boldsymbol{X})=\sum_{q=1}^{Q}p_q\log p_q \tag{13}$$

式中，p_q 表示随机变量 $\boldsymbol{X}$ 取值 q 的概率，而 $\boldsymbol{X}$ 可能的取值（状态）总数为 Q，当对数以 2 为底时，$H(\boldsymbol{X})$ 的单位为比特（bit），以下默认 2 为底。信息熵度量的是作为信源的随机系统的不确定程度。从公式(13)可知，当各个符号出现的机率相等，预测其对应的随机变量的取值的不确定性最高，信息熵最大（$\log Q$）。故信息熵可以视为不确定性的度量。熵也表示了需用于存储或传输消息中一个符号的平均比特数，如信源的熵。

经典的信息论多指通信系统，其模型是信源→信源编码→信道编码→信道传输＋噪声→信道解码→信源解码→信宿。信源编码是指，为了减少信源输出符号序列中的剩余度，提高符号的平均信息量，对信源输出的符号序列所施行的变换。其作用之一是设法减少码元数目和降低码元速率，即通常所说的数据压缩；作用之二是将信源的模拟信号转化成数字信号，以实现模拟信号的数字化传输。具体说，就是针对信源输出符号序列的统计特性来寻找某种方法，把信源输出符号序列变换为最短的码字序列，使后者的各码元所载荷的平均信息量最大，同时又能保证无失真地恢复原来的符号序列。一切旨在减少冗余度而对信源输出符号序列所施行的变换或处理，都可归入信源编码的范畴，如 CS。

Shannon 信源编码定理确立了数据压缩的可能达到的极限（在熵的意义下）。若独立同分布（i. i. d.）的随机变量数据流的长度趋近无穷大，数据压缩不可以使得代码信息率（单符号的平均比特数）少于信源的熵，否则造成信息丢失（但可以使代码信息率任意地接近信源熵）。数学描述如下：给定独立同分布的随机变量 $\boldsymbol{X}$，其时间序列 $X_1, X_2, \cdots, X_N$ 也是独立同分布，熵为 $H(X)$。对于任意 $\varepsilon^*>0$，对于任何大于 $H(X)$ 的信息率 R（编码后符号序列携带的信息量），存在一足够大的 N 和一编码器（生成 N 个独立同分布信源的符号 $X^{1:N}$），将其映射至 $N(H(X)+\varepsilon^*)$ 二进制比特数，使得信源符号 $X^{1:N}$ 可以高概率地（至少 $1-\varepsilon^*$）从这些二进制比特数恢复。

例如，无失真信源编码定理称，由 N 个符号组成的、每个符号的熵为 $H(X)$ 的无记忆平稳信源符号序列 $X_1,X_2,\cdots,X_N$，可用 M_N 个符号 Y_1，$Y_2,\cdots,Y_{M_N}$（每个符号有 Q 种可能值）进行定长编码。对任意 $\varepsilon>0$，为使译码差错小于 ε，要求编码后符号序列携带的信息量不小于编码前信源输出的信息熵：

$$(M_N/N)\log Q\geqslant H(X)+\varepsilon \tag{14}$$

CS 与信源编码均力图消除冗余，减少数据量，但它们存在诸多不同，如表 1 所示。

表 1 CS 与信源编码的对比

Tab. 1 Comparison Between CS and Source Coding

	压缩感知	信源编码
依据	信号的稀疏性 采样的不相关性	信源符号的统计特性 符号序列之间的相关性和信息冗余
	k-稀疏 $\boldsymbol{A}$ 列向量间的 线性相关性	概率分布信息熵 $H(\boldsymbol{X})$ 信息效率，冗余度
恢复条件	Spark	唯一可译码 无失真信源编码定理
编码方法	分组编码： $\boldsymbol{Y}=\boldsymbol{AX}$	分组编码，将信源序列映射到码表中的一个码字；香农编码等各种方法
解码方法	各种重构算法	各种解码方法
结果的表述	重构概率、各种范数意义下的误差率失真	失真度 率失真

3.2 CS 与信源及信源编码

信源编码是研究无噪声条件下的编码。CS 过程通常受噪声干扰，可以等效为无噪声的观测值通过有噪声的信道，可通过信道和信道容量等概念和方法对其进行分析[17]。

通信上信道容量是指在特定噪音标准下,信道的最大(理论上)传输率。有噪信道编码定理(也叫做 Shannon-Hartley 或 Shannon 定理)指出,尽管噪声会干扰通信信道,但还是有可能在信息传输速率小于信道容量的前提下,以任意低的错误概率传送数据信息。

假设一个有噪信道,信道容量为 C ,信息以速度 R 传送,如果

$$R < C \tag{15}$$

则存在一编码技术使接收端收到的错误达到任意小的数值。反之(即在传送速度超过信道容量的时),信息传输的可靠性无保证。

考虑噪声条件下 CS 模型:

$$\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{W} \tag{16}$$

它可看作是模型(2)叠加了加性噪声 $\boldsymbol{W}$(M 维向量)。我们可以讨论其与稀疏含噪信道模拟信号传输的等价性。

假设在信道编码里,线性分组码编码器输出的码字与输入之间的关系为 $\boldsymbol{Y}^* = \boldsymbol{A}^* \boldsymbol{X}^*$,其中 $\boldsymbol{X}^*$ 是输入 k 长的信息符号,$\boldsymbol{Y}^*$ 是输出的 M 长的码字符号,$\boldsymbol{A}^*$ 是编码矩阵(维度 $M\times k, M>k$)。输出码字在信道中传输,受噪声干扰后的信号可描述为式(16)的形式;此时,$\boldsymbol{W}$ 是噪声经抽样判决后的差错图案(一般可视为稀疏向量)。与编码矩阵 $\boldsymbol{A}^*$ 相关联的是校验矩阵 $\boldsymbol{G}$(维度$(M-k)\times M$),$\boldsymbol{G}$ 与 $\boldsymbol{A}^*$ 是对偶的,即 $\boldsymbol{GA}^*=0$。编码器(模型(2))输出的码字 $\boldsymbol{Y}^*$ 在 $\boldsymbol{G}$ 中投影为 0(模 2 和),即 $\boldsymbol{GY}^*=0$。可以将式(16)的稀疏含噪信道模拟编码转换为无噪 CS 过程:

$$\boldsymbol{S} = \boldsymbol{GY}^* = \boldsymbol{GW} \tag{17}$$

式中,$\boldsymbol{S}$ 是伴随式;式(17)的导出依据是前已述及的 $\boldsymbol{GA}^*=0$。于是,根据接收符号 $\boldsymbol{Y}$、校验矩阵 $\boldsymbol{G}$、差错图案 $\boldsymbol{W}$(稀疏性)等,可利用 CS 方法精确恢复 $\boldsymbol{W}$。将接受码字 $\boldsymbol{Y}$ 减去 $\boldsymbol{W}$,可得无差错码字 $\hat{\boldsymbol{Y}}^*$;对于系统码,取其前 k 个符号,即得信号 $\boldsymbol{X}$ 的估值 $\hat{\boldsymbol{X}}^*$[18]。

由此可见,CS 和信道编码有更相似的结构和编解码过程,它们的对比关系如表 2 所示。

表 2　CS 和信道编码之间的对比关系

Tab. 2　Comparison Between CS and Channel Coding

	压缩感知	信道编码
模型	$\boldsymbol{Y}=\boldsymbol{AX}+\boldsymbol{W}$	$\boldsymbol{Y}=\boldsymbol{A}^*\boldsymbol{X}^*+\boldsymbol{W}$
描述	观测矩阵 $\boldsymbol{A}$	编码矩阵 $\boldsymbol{A}^*$
编码器	随机构造的性能好	随机构造性能好
输入	N 长　k-稀疏	k 长
输出	M 长	M 长
冗余	$M-k$	$M-k$
恢复条件	信号的距离度量,RIP	信道编码定理,最小码距(最小汉明距离)
解码或稀疏近似	解求欠定方程组,找稀疏解 各种重构算法 最大似然法为最佳估计法	求解欠定方程组,找稀疏解 各种译码算法 最大后验概率法为最佳译码法
相关联矩阵	与观测矩阵不相关的稀疏基 $\boldsymbol{\Psi}$	校验矩阵 $\boldsymbol{G}$ 与生成矩阵 $\boldsymbol{A}^*$ 对偶,即 $\boldsymbol{GA}^*=0$

3.3　CS 与率失真

在噪声存在的情况下,需考虑未知信号的估计值与真实值之间的失真度,并尽可能将其控制在一定范围之内。信息率失真函数值是在信源和允许失真 D 固定的情况下,选择一种试验信道使信息传输率最小(求极小值)。这个极小值 $R(D)$ 是在信源给定情况下,接收端以满足失真要求而再现信源消息所必需获得的最少平均信息量[2]。

假定信源 X 的率失真函数为 $R(D)$,信道 $\boldsymbol{Y}=\boldsymbol{AX}+\boldsymbol{W}$ 的信道容量为 C,Sarvoham 等[19]将观测方程和重构算法分别看作传输信号的信道和还原信号的解码器 D_X,利用信道传输 M 个观测值能携带的最大信息量 $M\times C$,大于或等于描述信源(被观测信号 X)所需的最小比特数 $N\times R(D)$来限定至少观测数。

对于服从独立同分布(高斯分布)条件的信源,在其率失真函数可以确定的前提下,采样比 $\rho=M/N$应当满足:

$$\rho \geqslant \frac{2R(\xi(D_X))}{\log(1+\mathrm{SNR})} \tag{18}$$

式中,$\xi(D_X)$为失真度;SNR 为信噪比。

3.4　CS 与 Fano 不等式

信息论里,Fano 不等式将有噪信道的平均信息损失与误判概率联系起来,用于找到任一解码器的误差概率的下界。根据 Cover 和 Thomas[2]的论述,对于两个随机变量 $\boldsymbol{X}$、$\boldsymbol{Y}$,Fano 不等式可表示为:

$$H(\boldsymbol{X}\mid\boldsymbol{Y}) \leqslant H(\boldsymbol{X}\mid\hat{\boldsymbol{X}}) \leqslant H(P(e)) + P(e)\lg(|\chi|-1) \tag{19}$$

式中，χ 表示 X 的支撑；$H(X|Y)$ 是条件熵；$P(e)$ 是误差（随机事件 e：$\hat{X}\neq X$）概率；$H(P(e))$ 是二进制熵。简单的解释是，当且仅当条件熵 $H(X|Y)$ 为零（熵的最小值）时（即当 X 是 Y 的函数时），我们可以零误差地由 Y 估计 X，而当 $H(X|Y)$ 增加（降低）时，误差概率也将增加（降低）。

CS利用少量的观测值恢复稀疏信号。其中，寻找稀疏信号的支撑集是重要任务。观测值的个数影响信号重构的性能。利用费诺不等式的相关原理与结论可以推导出支撑集重构中的这一必要观测数。

针对近似支撑集重构，Reeves 和 Gastpar[20] 从信息论的角度推导出采样比的下限值。这一结论适用于任意的重构算法，但是未知信号 X（向量）及观测矩阵 A 需满足部分或全部限制条件。S^* 表示 X 的支撑集，$\hat{S}$ 是已知 Y（观测值向量）和 A 的前提下按照某重构算法对 S^* 的估计值，$d(S^*,\hat{S})$ 表示 $\hat{S}$ 和 S^* 之间的失真度。

假设 $S^*\rightarrow(Y,A)\rightarrow\hat{S}$ 构成 Markov 链，根据 Fano 不等式，可以推导出概率 $\Pr[d(S^*,\hat{S})>D]$ 与互信息 $I(S^*;Y,A)$ 之间的关系；使用互信息的链定理和数据处理定理限定 $I(S^*;Y,A)$，并依据信道编码定理，可得出采样比 ρ 的下限（与式(18)类似）：

$$\rho\geqslant\frac{2R(D;\kappa)}{\log(1+\mathrm{SNR})}\tag{20}$$

式中，SNR 是信噪比，$R(D,\kappa)$ 是率失真函数[20]。

3.5 CS算法的复杂度问题

首先定义 X 的 $\iota_p(p>0)$ 范数：$\|X\|_{\iota_p}=(\sum_{i=1}^{N}\|X_i\|^p)^{1/p}$。严格地说，$\iota_0$ 不是一个范数，$\|X\|_{\iota_0}=|i|X_i\neq0|=k$（如果 X 是 k-稀疏）。值得说明的是，ι_2 范数不合适。

首先，k-稀疏的信号 X 可以通过遍历搜寻（ι_0 最小搜寻）

$$\min_X\|X\|_{\iota_0}\quad \text{s. t.}\ \ AX=Y$$

所需的最小采样数为：

$$M\geqslant2k\tag{21}$$

其次，可以使用松弛算法，即 ι_1 最小化方法：

$$\min_X\|X\|_{\iota_1}\quad \text{s. t.}\quad AX=Y$$

所需的最小采样数为：

$$M\geqslant c_1k\log N\tag{22}$$

式中，c_1 为常数。两种解法（ι_0 最小搜寻与 ι_1 最小化）的解是等价的[14]。

ι_1 最小化算法能以接近1的概率精确恢复 X，其条件是观测数 M 为：

$$M\geqslant ck\lambda>2k\tag{23}$$

其中，λ 是过采样(over sampling)因子：对于高斯和贝努力观测矩阵而言，$\lambda=\log N$；而为 Fourier 变换矩阵[6]时 $\lambda=(\log N)^6$。

4 结语

综上所述，信息论已经在CS的发展进程中发挥着不可替代的作用。例如，信源、信道（信道容量）、率失真等概念直接应用于CS，信道编码定理和费诺不等式等信息论原理可用于确定CS的性能限（采样数的充要条件）。另一方面，CS的发展同时也丰富了信息论的内涵和外延。实际信息的稀疏性（使得低于Nyquist采样率的CS采样成为可能）、随机采样的信息感知性能（随机采样矩阵保持信息）、基于凸优化的信息重构方法（不同于Shannon-Nyquist采样中以Sinc函数为权重的线性内插信号重构方式）等CS机制，使其实现了数据采集与压缩的一体化。本文阐释了CS和信息论的联系，初步理清了CS及其sub-Nyquist采样率的信息论基础，有利于构建CS的信息论。

现有CS有关采样数的结论需要实践的检验。对于CS中常用的随机观测矩阵，现有文献已经给出有关这类矩阵相关性的上下限，然而对于在压缩感知雷达成像等应用领域[21-34]所涉及的卷积阵和一般性非随机观测矩阵的信息感知能力需要作进一步分析，并修正传统的基于随机观测矩阵的有关采样数的结论。由于CS应用经常涉及复信号或复图像[35]，对复信号的信源熵、信道容量等数量的计算较普通的实信号更为复杂，需要对此有通透的理解。此外，对相关应用之中系统参数的信息论解释也将有益于CS理论的完善。

参考文献

[1] Shannon C E. A Mathematical Theory of Communication[J]. *Bell System Technical Journal*, 1948, 27:379-423

[2] Cover T M, Thomas J A, Kieffer J. Elements of Information Theory[J]. *SIAM Review*, 1994, 36(3): 509-510

[3] Herman M A, Strohmer T. High-resolution Radar via Compressed Sensing[J]. *Signal Processing, IEEE Transactions on*, 2009, 57(6): 2 275-2 284

[4] Jiang Chenglong, Zhang Bingchen, Zhang Zhe, et al. Experimental Results and Analysis of Sparse Microwave Imaging from Spaceborne Radar Raw

Data[J]. *Science China Information Sciences*, 2012, 55(8): 1 801-1 815

[5] Candes E J, Tao T. Near-optimal Signal Recovery from Random Projections: Universal Encoding Strategies[J]. *Information Theory, IEEE Transactions on*, 2006, 52(12): 5 406-5 425

[6] Candès E J. Compressive Sampling[C]. Proceedings of the International Congress of Mathematicians, Madrid, 2006

[7] Donoho D L. Compressed Sensing[J]. *Information Theory, IEEE Transactions on*, 2006, 52(4): 1 289-1 306

[8] Baraniuk, R. Compressive Sensing. [J]. *Signal Processing Magazine, IEEE*, 2007, 24(4): 118-121

[9] Candès E J, Romberg J, Tao T. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information[J]. *Information Theory, IEEE Transactions on*, 2006, 52(2): 489-509

[10] Fletcher A K, Rangan S, Goyal V K. Necessary and Sufficient Conditions for Sparsity Pattern Recovery[J]. *Information Theory, IEEE Transactions on*, 2009, 55(12): 5 758-5 772

[11] Wainwright M J. Information-theoretic Limits on Sparsity Recovery in the High-dimensional and Noisy Setting[J]. *Information Theory, IEEE Transactions on*, 2009, 55(12): 5 728-5 741

[12] Aeron S, Saligrama V. Information Theoretic Bounds for Compressed Sensing[J]. *IEEE Transactions on Information Theory*, 2010 , 56(10): 5 111-5 130

[13] Akcakaya M, Tarokh V. Shannon-Theoretic Limits on Noise Compressive Sampling[J]. *IEEE Transactions on Information Theory*, 2010, 56(1): 492-504

[14] Candes E, Romberg J. Sparsity and Incoherence in Compressive Sampling [J]. *Inverse Problems*, 2007, 23(3): 969-985

[15] Peleg M, Shamai S. On Sparse Sensing of Coded Signals at Sub-Landau Sampling Rates[C]. Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of IEEE, Eilat, Israel, 2012

[16] Duarte M F, Eldar Y C. Structured Compressed Sensing: From Theory to Applications[J]. *Signal Processing, IEEE Transactions on*, 2011, 59(9): 4 053-4 085

[17] Aksoylar C, Atia G, Saligrama V. Sparse Signal Processing with Linear and Non-Linear Observations: A Unified Shannon Theoretic Approach [C]. 2013 IEEE Information Theory Workshop, Sevilla, 2013

[18] Li Linbo, Kang I. A Structured Construction of Optimal Measurement Matrix for Noiseless Sparse Recovery, Utilizing Channel Capacity and Polarization of Analog Transmission[OL]. http:arxiv.org/abs/1212.5577,2013

[19] Sarvotham S, Baron D, Baraniuk R G. Measurements vs. Bits: Compressed Sensing Meets Information Theory[C]. The 44th Annual Allerton Conference on Communication, Control and Computing, Illinois, USA, 2006

[20] Reeves G, Gastpar M. The Sampling Rate-distortion Tradeoff for Sparsity Pattern Recovery in Compressed Sensing[J]. *Information Theory, IEEE Transactions on*, 2012, 58(5): 3 065-3 092

[21] Cetin M, Karl W C. Feature-Enhanced Synthetic Aperture Radar Image Formation Based on NonquadraticRegularization[J]. *IEEE Transactions on Image Processing*, 2001, 10(4):623-631

[22] Baraniuk R, Steeghs P. Compressive Radar Imaging [C]. IEEE Radar Conference, Waltham, USA, 2007

[23] Alonso M T, Lopez-Dekker P, Mallorqui J. A Novel Strategy for Radar Imaging Based on Compressive Sensing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(12): 4 285-4 295

[24] Wei S J, Zhang X L, Shi J, et al. Sparse Reconstruction for SAR Imaging Based on Compressed Sensing [J]. *Progress in Electromagnetics Research*, 2010, 109: 63-81

[25] Potter L C, Ertin E, Parker J T, Cetin M. Sparsity and Compressed Sensing in Radar Imaging[J]. *Proceedings of the IEEE*, 2010, 98(6): 1 006-1 020

[26] Ender J H G. On Compressive Sensing Applied to Radar[J]. *Signal Processing*, 2010, 90: 1 402-1 414

[27] Jiang Hai, Zhang Bingchen, Lin Yueguan, et al. Random Noise SAR Based on Compressed Sensing [C]. IGARSS 2010, Hawaii, USA, 2010

[28] Patel V M, Easley G R, Healy D M, et al. Compressed Synthetic Aperture Radar[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2010, 4(2): 244-254

[29] Huan Yifeng, Wang Junfeng, Tan Zhen, et al. SAR Imaging Based on Compressed Sensing [C]. IGARSS 2011, Vancouver, Canada, 2011

[30] Zhang Bingchen, Hong Wen, Wu Yirong. Sparse Microwave Imaging: Principles and Applications [J]. *Science China Information Sciences*, 2012, 55(8):1 722-1 754

[31] Zhang Zhe, Zhang Bingchen, Hong Wen , et al.

Waveform design for L_q Regularization Based Radar Imaging and an Approach to Radar Imaging with Non-Moving Platform[C]. The 9th European Conference on Synthetic Aperture Radar, Nuremberg, Germany, 2012

[32] Yang Jungang, Thompson J, Huang Xiaotao, et al. Random-Frequency SAR Imaging Based on Compressed Sensing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(2): 983-994

[33] Ender J. A Brief Review of Compressive Sensing Applied to Radar[C]. The 14th International Radar Symposium IRS 2013,Dresden, Germany, 2013

[34] Zhang Jingxiong, Yang Ke, Guo Jianzhong. Information Theoretic Bounds for Compressed Sensing in SAR Imaging[C]. IOP Conf Ser : Earth Environ Sci ,Beijing,2014

[35] Yu Siwei, Khwaja A S, Ma Jianwei. Compressed Sensing of Complex-valued Data[J]. *Sinal Processing*, 2012, 92: 357-362

Information-theoretic Interpretations of Compressive Sampling

ZHANG Jingxiong[1] *YANG Ke*[1] *GUO Jianzhong*[2]

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

2 School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan 430074, China

Abstract: Compressive sampling or compressed sensing (CS) is a new paradigm for data acquisition and signal recovery. There are two-way relationships between CS and information theory: the former should and can be analyzed from the perspective of the latter, while the latter's content and extent are enriched and broadened by the former. Specifically, some basic concepts and theorems in information theory, such as source, channel, source coding, channel coding, rate distortion, Fano inequality, and the data processing theorem, provide theoretical foundation for research on CS, in particular, that concerning performance limits (e. g. , sampling rates). CS provides a highly efficient strategy for collecting, storing, transmitting, and reconstructing sparse signals through its unique concepts and algorithms, such as the sparsity of real signals (enabling CS sampling at a rate lower than Nyquist rate), the information sensing capacity of random sampling matrices (which preserve information); and information reconstruction based on convex optimization (different from signal reconstruction by Sinc kernels in the Shannon-Nyquist sampling theorem). Thus, CS is a mechanism for direct information sampling and processing, extending the domain of classic information theory. This paper seeks to clarify and explain the relationships between CS and information theory, revealing some of the fundamental issues in CS, in particular, those concerning CS sampling, and providing guidance for CS research directions.

Key words: compressive sampling; information theory; sparse; sampling; entropy

First author: ZHANG Jingxiong, PhD, Professor, specializes in fundamental and applied research on remote sensing and geography. E-mail: jxzhang@whu.edu.cn

Foundation support: The National Program on Key Basic Research Project (973 Program) of China, No. 2010CB731905; the National Natural Science Foundation of China, No. 41071286; the Research Program of Hubei Provincial Department of Education, No. D201416022.