

“People Hearing Without Listening:” An Introduction To Compressive Sampling 「不听即闻」——压缩采样介绍

Emmanuel J. Candès and Michael B. Wakin

1 介绍

传统信号或者图像采样多采取香农采样定理：即采样频率为信号频率最大值（那奎斯特采样频率）的二倍。

本文提出一种新的信号处理思路——压缩采样（CS）。在这种新方法下，采样率会大大降低。压缩采样依赖两条原则：稀疏性和不相干性，前者从属于信号，后者从属于感测模式。

- 稀疏性 (sparsity)：表达的概念是连续时间信号的信息采集率可能要比按带宽选择的采集率小得多，或者说离散时间信号依赖的自由度数要比他的长度小的多，更明确的说，压缩采样探索如下事实：许多自然信号在下述意义上是稀疏的或可压缩的：当用适当的基表示时，它们有更简洁的表达。
- 不相干性 (incoherence)：延伸了时间与频率之间的对偶性，它表达的概念是：正如时间域内的 Dirac 和尖峰信号在频率域内是展开的那样，具有以 Ψ 稀疏表示的对象，在获取它们的区域内一定是展开的。

极为重要的是人们可以设计出有效的传感或采样方案，捕捉嵌入在稀疏信号内有用的信息内容，并将其压缩成少量数据。这些方法不是自适应的，且只需要与少量的与提供信号简洁表述之基不相干（否则将失去测量过程对信号的依赖性）的固定波形相关联；此外，有利用数值优化由采集的少量数据重构全长信号的方法。换言之，CS 是一种非常简单有效的信息捕获方法，它以与信号无关的方式和低采样率采样，而后根据看似不完整的测量数据集用计算能力重构信号。

2 感知问题

对于信号 $f(x)$ ，传感机制为：

$$y_k = \langle f, \varphi_k \rangle, \quad k = 1, \dots, m. \quad (1)$$

也就是说我们只需要将要获得的对象与波形 φ_k 相关联，这是一个标准架构，例如，如果检测波形是 Dirac delta（尖峰）函数，那么 y 就是 f 在时间或空间一个采样值的矢量；如果检测波形是像素的指示函数，那么 y 就是数字相机中传感器特别采集的图像数据；最后，如果检测波形是正弦函数，那么 y 就是 Fourier 系数，磁共振成像 (MRI) 用的就是这种检测模态。

尽管可以建立起时空连续的 CS 理论，但这里只关心离散信号 $f \in \mathbb{R}^n$ ，原因有二，一是简单，二是现有理论在此情况下成熟得多。我们关心欠采样的情况，可得到的测量数 m 远小于信号 f

的维数 n 。由于各种原因，这种情形极为常见，譬如传感器数量有限，或者某些借助于中子散射获取图像的花费极为昂贵，或者是检测过程缓慢，象 MRI 那样只能对对象施行几次测量，等等。

这些情况产生一些重要问题，根据 $m \ll n$ 次测量能够准确的重构信号吗？能够设计出 $m \ll n$ 个检测波形捕获到几乎所有的关于 f 的信息吗？以及如何根据这些信息逼近 f ？毫无疑问，这些都是困难的事情，因为可能需要解欠定线性方程组。令 A 是以 $\varphi_1^*, \varphi_2^*, \varphi_3^*, \dots, \varphi_m^*$ 为行的 $m \times n$ 测量矩阵， φ^* 是 φ 的复转置。当 $m < n$ 时，由 $y = Af \in \mathbb{R}^m$ 复原 $f \in \mathbb{R}^n$ 的过程一般不是适定的，有无穷多组解，也许人们会根据具体问题想出一种好办法，这就是本文要讨论的问题。

3 稀疏性和不相干性

3.1 稀疏性

通过选择合适的基，许多自然信号都有简洁表示，考虑如图1的图像及其小波变换，尽管原始图像几乎所有的像素都是非零的，但小波系数提供的简明汇总是：绝大部分小波系数的值是小的，为数不多的大系数捕捉了有关对象的主要信息。

用数学的语言描述：我们有一个矢量 $f \in \mathbb{R}^n$ ，以正交基（譬如小波基） $\Psi = [\Psi_1, \Psi_1, \dots, \Psi_1]$ 展开如下：

$$f(t) = \sum_{i=1}^n x_i \psi_i(t) \quad (2)$$

x 是系数列， $x_i = \langle f, \Psi_i(t) \rangle$ ，将 f 表示成 \mathbf{x} 。现在明确稀疏性的含义：当信号有稀疏展开时，可以丢掉小系数而不会失真。按正式的说法， $f_s(t)$ 是保留展开式(2)中 S 个最大系数 (x_i) 值得到的结果。有定义 $f_s := \Psi x_s$ ，此后 x_s 就是系数向量 x_i ，只不过除了 S 个最大值外其余都是 0，该向量在严格意义上是稀疏的；因为，除了少数几个非零元素外，其余元素都为 0。

我们称这种几乎 S 个非零元的对象为 S -稀疏的，由于 Ψ 是正交基，我们有： $\|f - f_s\|_{\ell_2} = \|x - x_s\|_{\ell_2}$

如果 x 在按值排序快速衰减的意义上是稀疏的， x 就能很好地用 x_s 逼近，误差 $\|f - f_s\|_{\ell_2}$ 就是小量。通俗点说就是除了几个大系数外扔掉其余系数，不会造成太大的损失。从图1给出的例子，几乎觉察不到 1 兆像素图像与丢掉 97.5% 系数的近似图像之间的差别。

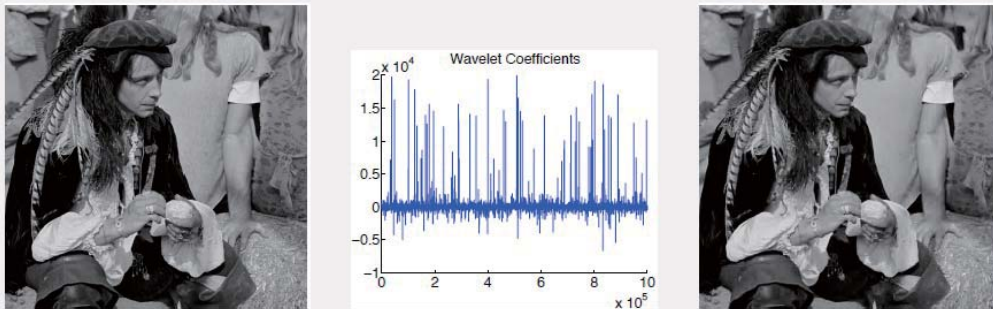


Fig. 1. Original megapixel image with pixel values in the range $[0,255]$ and its wavelet transform coefficients (arranged in a random order for enhanced visibility). Relatively few wavelet coefficients capture most of the signal energy. Many such images are highly compressible and the rightmost image shows the reconstruction obtained by zeroing out all the coefficients in the wavelet expansion but the 25,000 largest (pixel values below 0 and above 255 are set to 0 and 255 respectively). The difference with the original picture is hardly noticeable.

Figure 1:

这一原理为现代有损失编码基础，如 JPEG-2000 及其它编码器，由于有了一种数据压缩的简

单方法，将只需根据 f 计算 x ，然后（自适应地）将 S 个重要系数的值及位置进行编码。这种压缩方法需要知道所有 n 个系数 x ，由于事先不知道重要信息片段的位置（它们与信号有关，在我们的例子中有积聚在图像边缘的趋势，其位置事先未知）。在更一般的意义上，稀疏性是一种基本模型工具，它允许有效的基本信号处理，例如：准确的统计计算和分类，有效的数据压缩等等。本文要研究的事情是它更惊奇深远的含义，这就是信号稀疏性在数据采集本身所具有的重要作用。稀疏性决定着人们如何能够有效、非自适应的获取信号。

3.2 非相干采样

假定我们给定 \mathbb{R}^n 内一组正交基 (Φ, Ψ) ，第一个基 Φ 用于像式 (1) 那样感知对象 f ，第二个基用于表示 f ，对这一对基的限制不是实质性的，仅仅为了简化处理。

Definition. 1

Φ 和 Ψ 之间的相干度：

$$\mu(\Phi, \Psi) = \sqrt{x} \max_{1 \leq k, j \leq n} |\langle \varphi_k, \psi_j \rangle| \quad (3)$$

用通俗的话说，相干度度量着 Φ 和 Ψ 中任意两个元素的最大相关性。如果 Φ 和 Ψ 含有相干元，相干系数 μ 值就很大，否则值是少量。无论大和小要满足 $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$ 。上端点是因为基为单位向量，下端点是由于 Parseval 关系的结果，该关系指出，对每一个 j ， $\sum_{k=1}^n |\langle \varphi_k, \Psi_j \rangle|^2 = \|\Psi_j\|_{\ell_2}^2 = 1$ 。压缩采样主要与低相干度对有关，下面给出这样相干度对的一些例子。

第一个例子中 Φ 是经典的或尖峰基 $\varphi_k(t) = \delta(t-k)$ ， Ψ 是 Fourier 基， $\Psi_j(t) = n^{-1/2} e^{i2\pi jt/n}$ 。由于 Φ 是测量矩阵，这相应于时、空经典的采样格式，时间频率对遵循 $\mu(\Phi, \Psi) = 1$ ，因此有最大相干性。此外，尖峰信号与正弦波不仅在 1-D 情况下不相干，在任意维数都如此，2D、3D 等等相同。

第二个例子 Ψ 是小波基， Φ 取 Noiselets。Noiselets 同 Harr 小波基的相干度是 $\sqrt{2}$ ，同 Daubechies D4 和 D8 小波的相干度分别是 2.2 和 2.9，跨越非常大的样本容量 n 的范围；延伸至高维情况也是如此。（Noiselets 同尖峰信号也是最大不相干的，同 Fourier 基不相干）。由于以下事实我们对 Noiselets 感兴趣：(1) 它们与提供图像数据及其它数据稀疏表示的系统不相干；(2) 它们适合于非常快速的算法，Noiselets 变换耗时 $O(n)$ ，和 Fourier 变换一样 Noiselets 矩阵再作用于向量时不需存储，这对于有效的数值计算是极为重要的，没有这一点 CS 不可能很实用。

最后一个例子，随机矩阵基本上同任何固定的基 Ψ 都是不相干的。可以通过在单位球上独立均匀采样的 n 个向量正交化，均匀随机地选一个正交基 Φ ，那么，以高概率使 Φ 和 Ψ 相干度大约为 $\sqrt{2 \log n}$ 。通过扩充具有独立同分布元素的随机波形 $\varphi_k(t)$ 也将展示出与固定表示 Ψ 有非常低的相干性，例如 Gauss 型或 ± 1 二进制元。注意到这里有一个非常奇特的暗示：如果不相干系统测量是良好的，那么高效的机制应该获得同随机波形的关联，例如白噪声！

3.3 欠采样和稀疏信号重建

从理论上，我们本来测的是 f 的 n 个系数，但是我们只观测到它们的子集，并采集数据：

$$y_k = (f, \varphi_k), \quad k \in M \quad (4)$$

这里 $M \subset \{1, \dots, n\}$ 是基数 $m < n$ 的子集，用这一信息，我们决定由 ℓ_1 -范数 ($\|x\|_{\ell_1} := \sum_i |x_i|$) 极小化复原信号，所提出的重构 f^* 由 $f^* = \Psi x^*$ 给出； x^* 是下述凸优化问题的解：

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \quad s.t. \quad y_k = \langle \varphi_k, \Psi \tilde{x} \rangle, \quad \forall k \in M \quad (5)$$

即，在所有与数据相容的对象 $f^* = \Psi x^*$ 中挑选出最小 ℓ_1 -范数的系数序列。

用 ℓ_1 -范数作为稀疏提升函数可以追溯回几十年前。早先应用在反射地震学，从限带宽数据中寻找稀疏反射函数（用以指示地表下各层的重要变化）。但 ℓ_1 -极小化不是复原稀疏解的唯一方法，也提出了其它方法，如贪婪算法。

第一个结果断言：当 f 足够稀疏时，可以证明，籍 ℓ_1 -极小化复原信号是准确的。

Theorem. 1

固定 $f \in \mathbb{R}^n$ ，并假设基 Ψ 下系数序列 x 是 S -稀疏的，在 Φ 域上随机均匀地选择 m 次测量。那么，如果对某一正常数 C 有：

$$m \geq C\mu^2(\Phi, \Psi) \cdot S \cdot \log n \quad (6)$$

解(5)式的准确率是可以充分保证的。

这里做三点解释：

1. 相干性的作用是清晰的，相干度越小，需要的采的样就越少，因此，在前面一节中强调低相干度；
2. 通过测量任一组 m 系数（可能远小于信号的表面需要）人们都不会蒙受信号损失；如果 (Φ, Ψ) 等于或接近于 1，那么 $S \cdot \log n$ 次采样就足够了，而不是 n 次；
3. 在事先不知道 x 非零坐标个数、位置的条件下，信号 f 可通过凸泛函极小化根据压缩数据集复原，关于它们的幅值事先完全未知。

这个定理确实提出了一个非常具体的捕获方案，在不相干域采样不是自适应的，采集后执行线性规划。按这一方法，得到实质上的压缩信号，所需要的是将数据解压缩的解码器，这是 ℓ_1 -极小化所起的作用。

事实上这个非相干采样是早先谱稀疏信号采样结果的推广；也许正是谱稀疏采样引发了我们已经证明而且今天继续证明的 CS 发展。假设我们对超宽带采样感兴趣，而谱稀疏信号形为：

$$f(t) = \sum_{j=0}^{n-1} x_j e^{i2\pi jt/n}, t = 0, \dots, n-1,$$

此处虽然 n 很大，但非零分量 x_j 小于或等于 S （我们可以理解为非常小）。我们不知道哪些频率是主要的，也不知道主要频率集合上的振幅。由于主分量集不一定是顺序整数的子集，Nyquist/Shannon 理论几乎对我们毫无帮助（因为我们不能事先限制带宽，甚至认为所有 n 倍采样都需要）。就这个特例而言，定理 1(3.3)告诉人们可以根据 $S \cdot \log n$ 采样顺序，用任意未知容量 S 的频率支集重构信号；此外，这些采样不必是精心选择的，几乎这个容量里的任一样本集都有效；图2给出了一个说明性的例子。

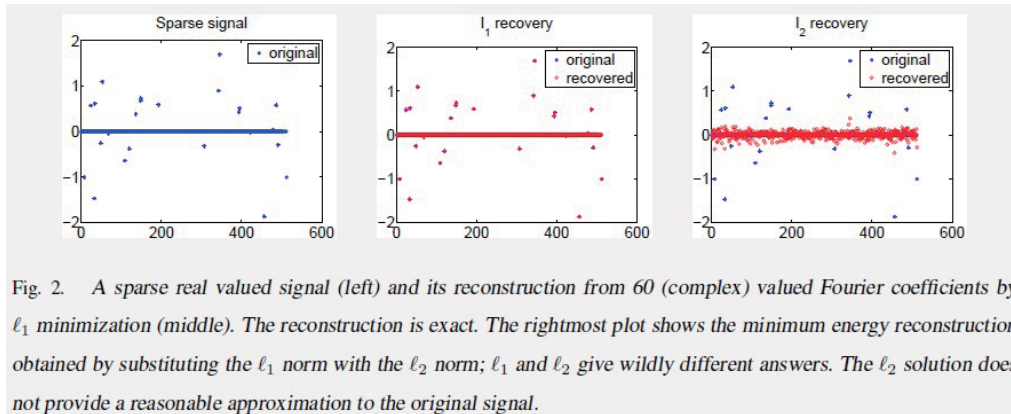


Fig. 2. A sparse real valued signal (left) and its reconstruction from 60 (complex) valued Fourier coefficients by ℓ_1 minimization (middle). The reconstruction is exact. The rightmost plot shows the minimum energy reconstruction obtained by substituting the ℓ_1 norm with the ℓ_2 norm; ℓ_1 and ℓ_2 give wildly different answers. The ℓ_2 solution does not provide a reasonable approximation to the original signal.

Figure 2:

现在讨论在这方面概率扮演的角色；人们需要借助概率阐述，因为人们不能希望容量 m 内的所有测量集都成立着类似结果。原因是有一些特殊的稀疏信号几乎在整个 Φ 域都为零，换言之，可以找到稀疏信号 f 和几乎等于 n 非常大容量的子集 (e.g. $n - S$)，对所有 $k \in M$ ， $y_k = \langle f, k \rangle = 0$ 。一方面给出这样的子集人们看到的尽是零，当然没有算法重构这种信号；另一方面，定理保证数据集中不发生准确复原的那一部分实际上是可忽略的；因此我们必须容忍极小的失败概率。对实用而言，如果采样容量足够大，失败的概率是零。

有趣的是，上述讨论的特殊稀疏信号至少也需要 $\mu^2 \cdot S \cdot \log n$ 个样本。用小的采样容量只是信息损失的概率太高，用任意多么复杂的方法重构都是不可能的。总的说，当相干度是 1 时，采样数不必大于 $S \cdot \log n$ ，但也不能再小。

我们以不相干采样的例子结束这一节。考虑图3所示演员的稀疏图像，与图1压缩的百万像素相同。研究对象是稀疏的，由于它的非零小波系数仅有 25000 个，然后取 96000 次非相干测量获取信息，并求解问题(5)。该例表明采样数大约为稀疏水平的 4 倍就满足了。许多研究者也报道了类似的成功经验，事实上有一个 4:1 的实用规则，该规则告知，为了准确的复原信号，每个未知非零项需要 4 个不相关采样。

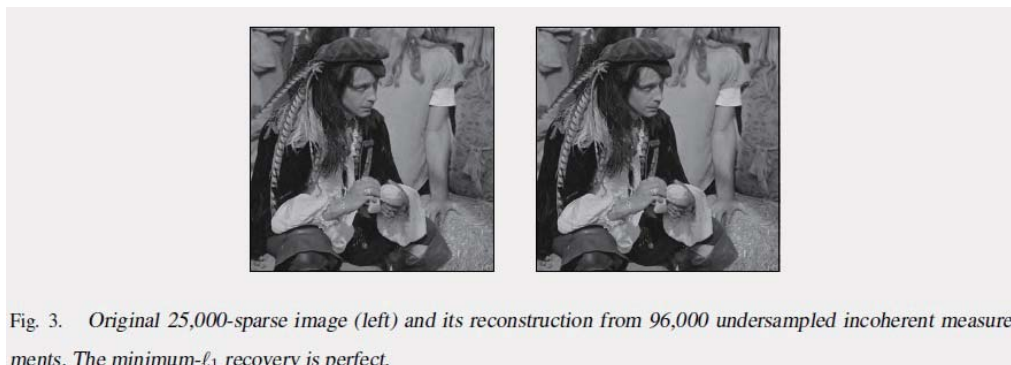


Fig. 3. Original 25,000-sparse image (left) and its reconstruction from 96,000 undersampled incoherent measurements. The minimum- ℓ_1 recovery is perfect.

Figure 3:

4 稳健压缩采样

我们已经证明可以仅由少量测量复原稀疏信号，但为了实际上更为有效起见，CS 需要能够处理带有噪声的近稀疏信号。

1. 首先，一般关心的对象并不是严格而是近似稀疏的，问题是：是否能够从高度欠采样测量中准确的复原这些对象。
2. 其次，在任何实际应用的测量数据中，由于测量设备不会有无限精度，总是存在哪怕是很小的噪声干扰。最轻微的情况下，噪声对数据的小扰动也会引起对重构的小扰动。

本节同时考察这两个问题。便于叙述，开始之前先考虑从形如

$$y = Ax + z \quad (7)$$

复原向量 $x \in \mathbb{R}^n$ ， A 是告诉我们 x 信息的 $m \times n$ 测量矩阵， z 是随机的或确定性的误差项。由于 $f = \Psi x$ ， $y = R\Phi f$ (R 是从 M 中提取采样坐标的 $m \times n$ 矩阵)，令 $A = R\Phi\Psi$ ，可以写成 $y = Ax$ ，因此人们可以用抽象模型(7)。 x 可以是以适当的基表示对象的系数序列。

4.1 约束等距性

约束等距性：restricted isometry property (RIP)，是对 CS 的一般鲁棒性非常有用的一个重要概念。

Definition. 2

对每一个整数 $s = 1, 2, \dots$ ，定义矩阵 A 的等距常数 δ_s 为使下式：

$$(1 - \delta_s) \|x\|_{\ell_2}^2 \leq \|Ax\|_{\ell_2}^2 \leq (1 + \delta_s) \|x\|_{\ell_2}^2 \quad (8)$$

对所有 s -稀疏向量 x 成立的最小值。

如果 δ_s 不太接近于 1，就不十分严谨的说矩阵 A 遵循 s 阶约束等距性。如果这一性质成立， A 近似保持 S -稀疏信号的欧几里得长度，反过来说 S -稀疏向量不可能在 A 的零空间（这是有用的，若不其然就没有重构信号的希望）。约束等距性的等价表述是，取自 A 的 S 列子集事实上近乎正交（ A 的列不可能准确正交，因为行大于列）。

为看到约束等距与 CS 之间的联系，想象一下我们用 A 获取 S -稀疏信号。首先假设 $\delta_{2S} < 1$ ，我们就声称能由数据 $y = Ax$ 复原信号 x ，事实上 x 是方程组 $y = Ax$ 的唯一稀疏解，即非零元的个数最少。为看清这一点，考虑任意其它解 $x + h, h \neq 0$ ，那么 $Ah = 0$ ， h 一定至少有 $2S + 1$ 个非零元素，则 $x + h$ 至少有 $S + 1$ 个非零元。相反，假设 $\delta_{2S} = 1$ ，那么， A 的 $2S$ 列必定线性相关，在此情形下 $2S$ -稀疏向量 h 遵循 $Ah = 0$ ，可以将 h 分解为 $x - x'$ 使 x 和 x' 都是 S -稀疏信号，由此得出 $Ax - Ax'$ ，这意味着我们找到了给出相同测量的两个稀疏向量，显然人们不能重构这样的稀疏信号。由此可见，要复原 S -稀疏信号，必须满足 $\delta_{2S} < 1$ ，或至少(8)中 $\|Ax\|_{\ell_2}$ 取低限。

4.2 欠采样复原信号

如果满足约束等距性，解线性规划问题：

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \quad s.t. \quad A\tilde{x} = y (= Ax) \quad (9)$$

得到的重构是准确的。

Theorem. 2

假设 $\delta_{2S} < \sqrt{2} - 1$ ，那么问题(9)的解 x^* 对某一常数 C_0 遵循：

$$\|x^* - x\|_{\ell_2} \leq C_0 \cdot \|x - x_S\|_{\ell_1} / \sqrt{S} \quad \text{and} \quad \|x^* - x\|_{\ell_1} \leq C_0 \cdot \|x - x_S\|_{\ell_1} \quad (10)$$

x_S 是除去最大的 S 个值其余置为零分量的向量 x 。这个定理的结论比定理 1(3.3)强。若 x 是 S -稀疏的, $x = x_S$, 信号复原是准确的; 若 x 不是 S -稀疏的, (10)式断言: 复原信号的质量恰如人们提前知道最大 S 个最大 x 值并确定直接测量的结果那么好。换言之, 重构近乎对对象完全了解的预言, 为我们抽取出 S 个最重要的信息片段。

与我们早先的结果的另一个惊人的差别在于它是确定性的, 不涉及到概率。如果我们足够幸运使测量矩阵 A 遵循定理的假设, 就可以用它, 并保证准确地复原所有 S -稀疏信号, 而对其它情况实质上是复原所有矢量的 S 个最大元素; 即, 没有失败的概率。此外, 相同测量矩阵对 \mathbb{R}^n 中所有向量都适用, 常称为通用方法。

用此方法我们失去的是遵循假设的 S (能有效复原的分量数) 与 m (测量数或矩阵的行) 之间的关系, 为了导出更有效的结果, 我们宁愿寻找 S 接近 m 满足约束等距性的测量矩阵。能设计出这样的矩阵吗? 我们下一节将证明这是可能的, 但首先考察一下面对数据干扰 CS 的鲁棒性。

4.3 噪声信号的稳健恢复

给我们如(7)所示的有噪声数据, 用 ℓ_1 极小化重构约束:

$$\min \|\tilde{x}\|_{\ell_1} \quad s.t. \quad \|A\tilde{x} - y\|_{\ell_2} \leq \epsilon \quad (11)$$

ϵ 为数据噪声值的界, 问题(11)在文献 [21]、[22] 中称为 LASSO; 就我们所知这首先是在 [8] 中提出的。这又是一个特有的凸优化问题, 准确的说是二阶圆锥规划, 求解此问题有多种有效的算法。

Theorem. 3

假设 $\delta_{2S} < \sqrt{2} - 1$, 那么问题(11)的解 x^* 对某一常数 C_0 和 C_1 遵循:

$$\|x^* - x\|_{\ell_2} \leq C_0 \cdot \|x - x_S\|_{\ell_1} / \sqrt{S} + C_1 \cdot \epsilon \quad (12)$$

这几乎不能更简单了, 重构误差有两项之和限定, 第一项是没有噪声时的误差, 第二项正比于噪声水平。此外, C_0 和 C_1 是特征小量, 例如在 $\delta_{2S} = 1/4$ 的情况下, $C_0 \leq 5.5$ 和 $C_1 \leq 6$ 。图4显示有噪声数据的重构。

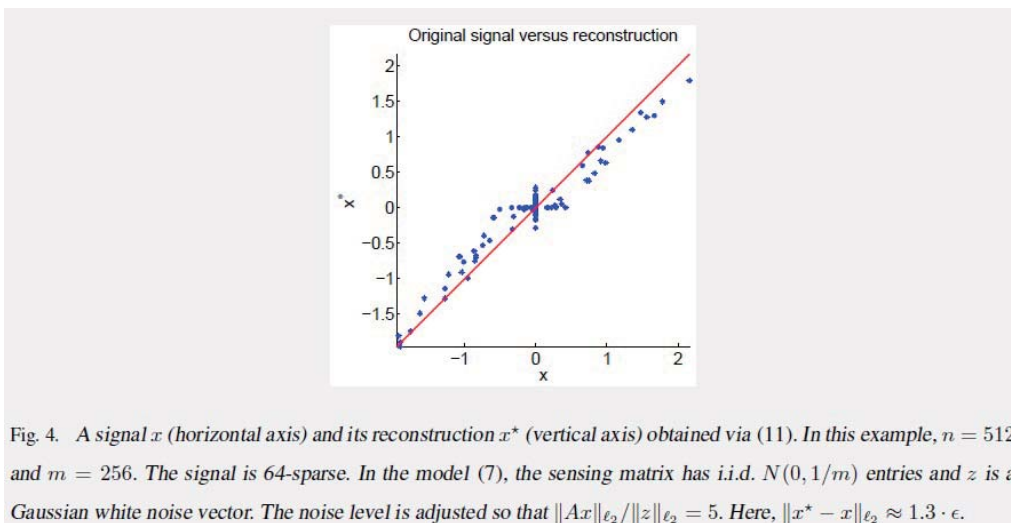


Figure 4:

最后这个结果为 CS 建立了实用有效的传感机制。它对各种不一定是稀疏的信号都有效, 在传

感方面能优雅地处理噪声。剩下要做的事情就是设计满足约束等距性 (RIP) 的测量矩阵; 这是下一节的目标。

5 随机测量

回到 RIP 的定义, 我们要寻找测量矩阵具有列矢量的任意子集都几乎正交的性质; 这个子集越大越好; 为此, 随机性重新出现。

考虑如下测量矩阵:

1. 通过在单位球 \mathbb{R}^m 上随机均匀地采集 n 列矢量, 形成 A 。
2. 通过从平均值为 0, 方差为 $1/m$ 的正态分布上独立同分布 (i.i.d.) 采集元素, 形成 A 。
3. 如节 3.2 所述对随机投影 P 采样并标准化为: $A = \sqrt{\frac{n}{m}}P$, 形成 A 。
4. 从对称 Bernoulli 分布 ($P(A_{i,j} = \pm 1/\sqrt{m}) = \frac{1}{2}$) 或其它亚 Gauss 分布上独立同分布采集元素, 形成 A 。

假如:

$$m \geq C \cdot S \log(n/S) \quad (13)$$

那么, 所有这些矩阵将以极大概率符合约束等距性 (即定理的条件), 上式中 C 为与具体情况有关的某一常数。命题 1 ~ 3 用了概率论中非常标准的结果。对命题 4 的论证更复杂些, 见 [23] 和 Pajor 及其合作者的研究, 如 [24]。

在所有这些例子中, 测量矩阵满足(13)式而又不具有约束等距性的概率是关于 m 指数型小量。有趣的是, 没有测量矩阵和重构算法用比式(13)左端少得多的采样而又能给出定理 2(4.2)的结果。在此意义上, 用上述测量矩阵及 ℓ_1 极小化是接近最优的感知策略。

可以像第三节中所作的那样建立正交基对的约束等距性, 用 $A = R\Phi\Psi$, R 随机均匀地抽取 m 个坐标, 使

$$m \geq C \cdot S (\log n)^4 \quad (14)$$

是以极大概率具有该性质的充分条件。如果要失败的概率对某 $\beta > 0$ 不大于 $O(n^{-\beta})$, 知(14)中的指数是 5 而不是 4 (人们相信(14)对恰为 $\log n$ 也是成立的)。这证明可以稳定而准确地由不相干域中戏剧性的欠采样数据重构近稀疏信号。

最后, 矩阵 $A = \Phi\Psi$ 也可以成立约束等距性, 这里 Ψ 是任意正交基, Φ 是从适当分布中随机抽取的测量矩阵。如果固定基 Ψ , 按命题 1 ~ 4 构成测量矩阵, 假若:

$$m \geq C \cdot \log(n/S) \quad (15)$$

那么, $A = \Phi\Psi$ 将以极大的概率符合约束等距性 (RIP); 式中 C 为与具体情况有关的某一常数。这种随机测量矩阵 Φ 在某种意义上是普适的, 当设计测量系统时甚至不需知道稀疏基。

6 什么是压缩采样?

数据获取通常按下述方式工作: 大量的数据只是采集出来, 其中一大部分在压缩阶段被抛弃, 为了储存和传送通常必须这样做。用本文的话说, 人们获得高分辨率的像素数组 f , 计算完整的一套变换系数, 将最大的系数编码, 丢掉其它系数, 实质上以 f_S 结束; 这种大规模采集数据然后压缩的处理方法是极为浪费的 (可以想象一下数字相机有百万像素的图像传感器, 而最终编码的图像只有几百 k 字节)。

CS 的处理极为不同，其表现好像是能够直接获取处理对象恰为最重要的信息。通过取如第 5 节的 $O(S \log(n/S))$ 个随机投影，人们就有足够的信息重构信号，其精度至少达到 f_S 具有的精度，得到处理对象的最佳 S 项逼近和最优压缩表示。换言之，CS 数据获取协议实质上是把模拟信号转换成压缩后的数字形式，使得人们至少从原理上可以从仅有为数不多的传感器得到超分辨信号。采集步骤后所有需要做的就是将测量数据解压缩。料想不到的是采集步是固定的，尤其是根本不试图理解信号的结构，好像是「不听即闻」(hearing without listening)。

CS 同编码理论，更明确的说同 Reed-Solomon 编码的理论和实践有某些表面上的相似性。就本文讨论内容简单地讲，如所周知，可以采用编码理论的概念如下建立 CS：人们可以根据信号的前 $2S$ 个 Fourier 系数

$$y_k = \sum_{t=0}^{n-1} x_t e^{-i2\pi kt/n}, \quad k = 0, 1, 2, \dots, 2S - 1 \quad (16)$$

或者从相继 $2S$ 个信号频率集唯一的重构任何 S -稀疏信号（复原信号的计算成本实质上是解 $S \times S$ Toeplitz 矩阵，或计算 n 点 FFT），这意味着可以用这一方法测量可压缩信号吗？由于两个原因答案是否定的。

其一，Reed-Solomon 解码是代数方法，不能用于非稀疏信号（解码是通过求多项式的根寻找信号支撑集）；

其二，根据信号的前 $2S$ 个 Fourier 系数寻找信号支撑集的问题，甚至当信号具备准确稀疏性时，是非常不适定的（这个问题与由少量值高度集中的数预测高阶多项式相同）；哪怕系数受到微小的扰动也会导致完全不同的答案，这使得用有限精度的数据可靠地计算支撑集在实际上是不可能的。反之，纯粹代数方法忽略了信息算子的调节作用，使取得良态矩阵，这是精确估计的关键，正如约束等距性所起作用为 CS 所关心的中心问题。

7 应用

可压缩信号可以用与信息水平 $S \ll n$ 成正比的若干不相干测量捕获，这一事实有着深远的意义并涉及到许多可能的应用。

1. **数据压缩**：对数据压缩而言，在某些情况下解码器上稀疏基可能是未知的，或者施行起来不实用。然而正如第五节所讨论的随机设计的 Φ 可视为普适解码策略，因为它不需要围绕 Ψ 的结构设计（只有解码或复原 f 时才需具备 Ψ 的知识和实现 Ψ 的能力）。这种普适性特别有助于诸如传感器网络之类的多信号装置的编码。关于这个问题请读者参考 Nowak 等人及 Goyal 在其它地方的论文。
2. **信道编码**：正如文献 [15] 所解释的那样，可以围绕 CS 的原理（稀疏性，随机性，和凸优化）并将其用于设计快速纠错码，避免错误信号的传输。
3. **反问题**：对于其它一些情况，获取 f 的唯一方法可能是用某种模态的测量系统 Φ ，假定与 Φ 不相干 f 的稀疏基 Ψ 存在。就有可能进行有效的测量，一个应用涉及到 MR 血管造影术和其它类型的 MR 设备；在这些例子中 Φ 记录 f 的 Fourier 变换的子集，所希望的图像 f 在时间域和小波域都是稀疏的。其它领域的这个问题 Lustig 等人有更深入的讨论。
4. **数据采集**：有时全部测量模拟信号的 n 个离散时间样本可能难以得到（而且接下来也难以压缩）。在这种情况下，可能有助于设计出物理采样设备，直接记录传入模拟信号离散的、低采样率的不相干测量。

最后一个应用暗示我们，数学和计算方法可能会对传统的硬件设计有限制的领域产生巨大的冲击。如传统的图像设备采用 CCD 和 CMOS 技术基本上限于可视光谱，而 CS 相机用数字微型镜

头阵列采集不相干数据，也许能够明显扩充其能力。关于这个问题其它地方的讨论，以 Baraniuk 讨论这种相机更加深入。

有一部分研究者已专心于大带宽高级模拟-信息 (A/I) 转换设备的研究，目标是帮助减轻传统模拟-数字 (A/D) 转换技术的压力，当前该技术限于 1GHz 的采样速度。作为可选择的方案，我们建议两种特殊的、可以从高带宽模拟信号中获得离散的、低采样率，不相干测量序列的 A/I 架构。对高阶逼近，每一个测量 y_k ，都可以解释为入射模拟信号 f 与模拟测量波形 φ_k 的内积 $\langle f, \varphi_k \rangle$ ，正如在离散 CS 框架那样，我们的初步结果表明遵循稀疏或可压缩模型（在某一模拟字典 Ψ 中）的信号可以用这些设备以正比于信息水平而不是那奎斯特频率有效的捕获。当然，利用离散 CS 方法复原模拟稀疏信号有待克服的挑战性问题。全面解决这些问题超出了本短文的范围；人们可以简单的接受其思想，离散/采样稀疏字典允许适当的复原。

我们有两个架构如下：

1. **非均匀采样器 (NUS)**：此架构只是把随机采样时间点上的信号数字化，即 $y_k = f(t_k) = \langle f, \delta_{t_k} \rangle$ ，事实上这些随机或伪随机时间点是通过对抖动规则格子上的名义（低采样率）采样点得到的。由于尖峰与正弦信号不相干，因此这种架构可用于对具有远低于那奎斯特频率的稀疏频谱信号采样。当然与减少采样率有关的好处是极大的，因为这提供了附加电路稳定时间，并具有减小噪声水平的作用。
2. **随机预积分 (RPI)**：此架构可用于种类更广泛的稀疏域，在时间-频率平面有稀疏特征的最显著的那些信号。鉴于没有可能以极高的采样率将模拟信号数字化，而极有可能以很高的速率改变它的极性。RPI 架构的思想（见图5）就是以正 1 和负 1 的伪随机序列乘以信号，将乘积在时间窗口积分，将时段末的积分值数字化，这是一个并行架构，有若干个这样的乘法器-积分器对用完全不同的甚至几乎独立的随机符号序列并列运行。事实上 RPI 架构将信号同 ± 1 符号序列库相关联，是普适的 CS 测量方法之一。

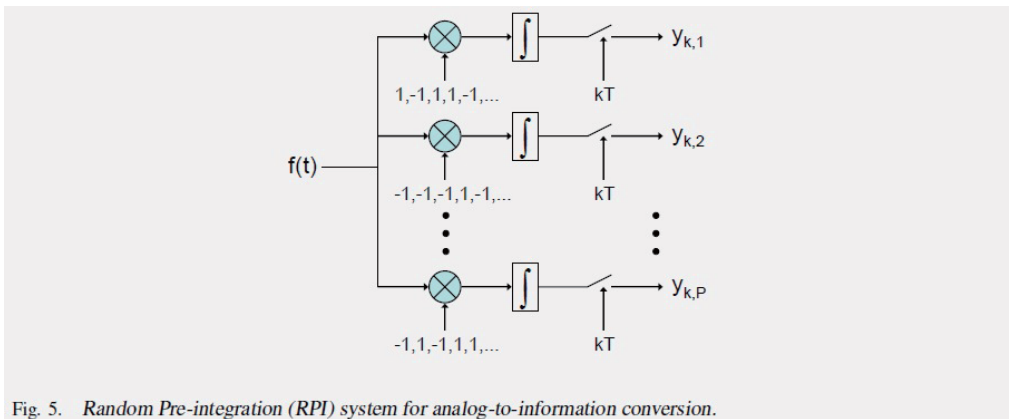


Fig. 5. Random Pre-integration (RPI) system for analog-to-information conversion.

Figure 5:

对上述每个架构，我们都曾用数值的方法确认，有些还用物理的方法证实，系统对于电路非理想，如热噪声、时钟误差、干扰及放大器非线性等情况是鲁棒性的。

将 A/I 架构应用于实际采集方案还需要发展 CS 算法和理论，包括研究模拟信号稀疏表示的字典。我们最后以一个离散的例子结束。对这个实验，我们取 f 是长度 $n = 512$ 的 1-D 信号，包含两个调制脉冲（图6左侧），从这个信号中我们用以独立同分布柏努利 ± 1 元素填充生成的 $m \times n$ 测量矩阵 Φ ，采集 $m = 30$ 次测量，这是不合理的小数据量，欠采样因子超过 17。为了重构信号，我们考虑时-频 Gabor 字典 Ψ ，它由 Gaussian 窗限时并具有不同位置和尺度的各种正弦波构

成。总的说字典是过完备 ($43 \times$ overcomplete) 的, 不含包括 f 的两个脉冲。图6中间图形表示取 $\|x\|_{\ell_1}$ 极小化使 $y = \Phi\Psi x$, 重构效果显示出明显的人为效应, 我们看到 $\|f - f^*\|_{\ell_2} / \|f\|_{\ell_2} \approx 0.67$, 而事实上我们通过将两者都变为 ℓ_1 复原方法, 就可以消除人为效应。首先, 我们改用极小化 $\|\Psi * \tilde{f}\|_{\ell_1}$ s.t. $y = \Phi\tilde{f}$; (当 Ψ 是正交基时这一改变不起作用); 其次, 当得到估计值 f^* 时我们将 ℓ_1 -范数重新加权, 重复重构过程, 对那些估计较大的系数采用较低的惩罚值; 图6右图显示了重复加权 4 次迭代的结果, 我们看到 $\|f - f^*\|_{\ell_2} / \|f\|_{\ell_2} \approx 0.022$ 。有关这个新方向的更多信息, 建议读者参阅文献 [30]。在此, 给出的要点是, 即便这个数据量小到荒谬的地步, 人们仍可以捕获到信号里包含的绝大部分信息。这就是 CS 在当前以及未来应用极有前途的原因。

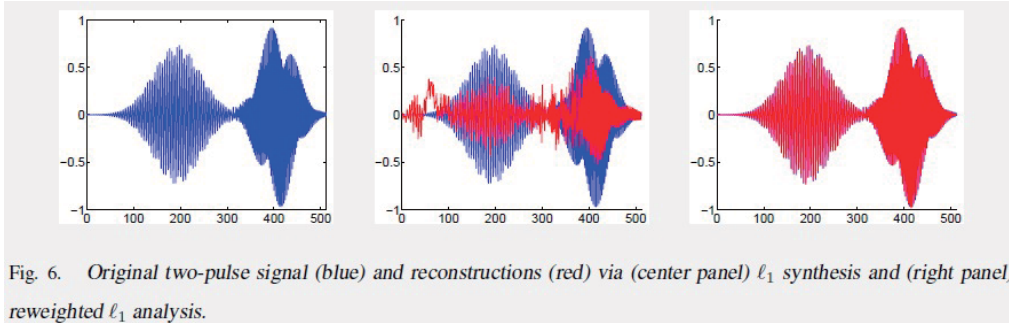


Fig. 6. Original two-pulse signal (blue) and reconstructions (red) via (center panel) ℓ_1 synthesis and (right panel) reweighted ℓ_1 analysis.

Figure 6:

References

- [1] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.
- [2] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006.
- [3] D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
- [4] A. Bilgin, M.W. Marcellin, and M.I. Altbach. Compression of electrocardiogram signals using JPEG2000. *Consumer Electronics, IEEE Transactions on*, 49(4):833–840, 2003.
- [5] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 47(7):2845–2862, 2001.
- [6] R. Coifman, F. Geshwind, and Y. Meyer. Noiselets. *Applied and Computational Harmonic Analysis*, 10(1):27–44, 2001.
- [7] J.F. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38:826, 1973.
- [8] F. Santosa and W.W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7:1307, 1986.
- [9] J. Tropp and A.C. Gilbert. Signal recovery from partial information via orthogonal matching pursuit, 2005.

- [10] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23:969, 2007.
- [11] A.C. Gilbert, S. Muthukrishnan, and M. Strauss. Improved time bounds for near-optimal sparse Fourier representations. In *Proceedings of SPIE*, volume 5914, page 59141A. Citeseer, 2005.
- [12] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *Signal Processing, IEEE Transactions on*, 50(6):1417–1428, 2002.
- [13] D.L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49(3):906–931, 1989.
- [14] P. Feng and Y. Bresler. Spectrum-blind minimum-rate sampling and reconstruction of multi-band signals. In *icassp*, pages 1688–1691. IEEE, 1996.
- [15] E.J. Candès and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.
- [16] E.J. Candès, J.K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.
- [17] EJ Candès. Lectures on compressive sampling and frontiers in signal processing. *The Institute for Mathematics and its Applications*, pages 2006–2007.
- [18] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *American Mathematical Society*, 22(1):211–231, 2009.
- [19] E. Candès and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351, 2007.
- [20] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *Information Theory, IEEE Transactions on*, 52(9):4036–4048, 2006.
- [21] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.
- [22] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999.
- [23] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.
- [24] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constructive Approximation*, 28(3):277–289, 2008.
- [25] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.
- [26] R.E. Blahut. *Algebraic codes for data transmission*. Cambridge Univ Pr, 2003.
- [27] D. Baron, M.B. Wakin, M.F. Duarte, S. Sarvotham, and R.G. Baraniuk. Distributed compressed sensing. *preprint*, pages 1–50, 2005.

-
- [28] M. Lustig, D.L. Donoho, and J.M. Pauly. Rapid MR imaging with compressed sensing and randomly under-sampled 3DFT trajectories. In *Proc. 14th Ann. Meeting ISMRM*. Citeseer.
- [29] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, KF Kelly, and RG Baraniuk. A compressed sensing camera: New theory and an implementation using digital micromirrors. *Proc. Computational Imaging IV at SPIE Electronic Imaging, San Jose*, 2006.
- [30] E.J. Candes, M.B. Wakin, and S.P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.