“压缩传感”引论

香港大学电机电子工程学系

高效计算方法研究小组

沙威　2008 年 11 月 20 日

Email: wsha@eee.hku.hk

Personal Website: http://www.eee.hku.hk/~wsha

到了香港做博士后，我的研究领域继续锁定计算电磁学。我远离小波和计算时谐分析有很长的时间了。尽管如此，我仍然陆续收到一些年轻学者和工程人员的来信，和我探讨有关的问题。对这个领域，我在理论上几乎有很少的贡献，唯一可以拿出手的就是几篇网上发布的帖子和一些简单的入门级的代码。但是，从小波和其相关的理论学习中，我真正懂得了一些有趣的知识，并获益良多。我深切地感到越来越多的学者和工程师开始使用这个工具解决实际的问题，我也发现互联网上关于这方面的话题多了起来。但是，我们永远不能只停留在某个阶段，因为当今学术界的知识更新实在太快。就像我们学习了一代小波，就要学习二代小波；学习了二代小波，就要继续学习方向性小波（X-let）。我也是在某个特殊的巧合下不断地学习某方面的知识。就像最近，我的一个友人让我帮她看看“压缩传感”（Compressive Sensing）这个话题的时候，我的兴趣又一次来了。我花了一个星期，阅读文献、思考问题、编程序、直到写出今天的帖子。我希望这篇帖子，能对那些没进入且迫切想进入这个领域的学者和工程师有所帮助。并且，我也希望和我一个星期前一样，对这个信号处理学界的“一个大想法”（A Big Idea）丝毫不了解的人，可以尝试去了解它。我更希望，大家可以和我探讨这个问题，因为我到现在甚至不完全确定我对压缩传感的某些观点是否正确，尽管我的简单的不到 50 行的代码工作良好。在这个领域中，华裔数学家陶哲轩和斯坦福大学的统计

学家 David Donoho 教授做出了重要的贡献。在这个引言中，我用简单的关键字，给出我们为什么需要了解甚至是研究这个领域的原因。那是因为，我们从中可以学习到，下面的这些：矩阵分析、统计概率论、拓扑几何、优化与运筹学、泛函分析、时谐分析（傅里叶变换、小波变换、方向性小波变换、框架小波）、信号处理、图像处理等等。所以，我们有什么理由，拒绝这个有意思的东西呢？让我们开始吧。

传统思路——正交变换

对于一维的信号 $x \in R^{N\times 1}$，大多数情况下，信息是冗余的。我们可以通过正交变换的方法来压缩它。正变换：$y = \Psi x$，反变换 $x = \Psi^H y$。这里，$\Psi\Psi^H = \Psi^H\Psi = I$，$\Psi \in C^{N\times N}$，$I$ 是单位矩阵。对于 $y \in C^{N\times 1}$，能量较 x 集中，本质上去除了 x 中的相关性。因此，我们只保留 K 个较大的分量，而把其它 $N-K$ 个置为零。通过反变换，我们能够近乎完美的重建原始信号。因为，那 $N-K$ 个变换域系数的贡献，实在微乎其微。具有这样性质的信号被称为 K “稀疏”（Sparsity）的。于是，我们有了如下编码解码的策略：

编码：构造 Ψ，做正变换 $y = \Psi x$，保留 y 中最重要的 K 个分量，和其对应的位置。

解码：把 K 个分量放回到对应的位置，其它位置填 0，构造 Ψ^H，反变换 $\hat{x} = \Psi^H \hat{y}$。

而解码能否近乎得到原始信号呢？显然，我们希望 $\| x-\hat{x} \|_2 = \| y-\hat{y} \|_2 \le \delta$，$\delta$ 是一个小的常数。但更有效的是用相对误差 $\| y-\hat{y} \|_2 / \| y \|_2 \le \delta$。

但这种编码解码方法有些缺点：1、考虑到香农（Shannon）采样定理，为了获得很好的信号分辨率，采样间隔会很小，造成了原始信号长度会很长，因此变换过程会消耗很长的时间。2、K 个需要保留的重要分量的位置，是随着信号的不同而不同的。因此，这种策略是“自适应”（Adaptive）的，且需要分配多余的空间存储这些位置。3、一旦在传输过程中 K

个分量中的某几个丢失了，后果可想而知。如果我们制作一个音频设备，1 将带来电力的消耗和用户的不满，2 将带来存储空间的增加，3 将带来较差的抗干扰能力。

新的思路——压缩传感

压缩传感（Compressive Sensing）是一个很有意思的新的方向。它也正成为信号处理领域的“A Big Idea”。对于信号 $x \in R^{N\times 1}$，我们可以找到它的 M 个线性测量（Liner Measurement），$s=\Phi x$。$\Phi \in R^{M\times N}$。这里，Φ 的每一行可以看作是一个传感器（Sensor），它与信号相乘，拾取（Acquisition）了信号的一部分信息。拥有了这 M 个测量和 Φ，我们就可以近乎完美的重构原始信号了。听起来“相当”传奇，事实上，它基于如下严格的数学最优化（Optimization）问题：

目标函数 $\min \|\hat{y}\|_0$，且满足等式约束 $\Phi\Psi^H\hat{y}=s$

或者，可以写成

$$\min \|s-\Phi\Psi^H\hat{y}\|_2+\lambda\|\hat{y}\|_0$$

求解该最优化问题，得到变换域的 $\hat{y}$，然后反变换，便可以得到时域的 $\hat{x}$。公式中的 2 是我们熟悉的 2-范数，而 0 是什么呢？是 0-范数，也就是向量 $\hat{y}$ 中非零元素的个数。看起来很有道理，因为 $\hat{y}$ 是待求的变换域向量，它是 K 稀疏的。使 $\hat{y}$ 非零元素的个数尽量小，也就是保留了尽量少的重要的 K 个分量，显然这几个分量可以近乎完美重构 x。我们回到传统的思路，这 K 个分量是我们在变换域“自适应”找的，而该优化算法也可以使我们找到这 K 个分量。

这就足够了吗？显然不行，我们仍然没有探讨测量矩阵 Φ 需要满足的性质。我们用极限分析法。如果我们把 Φ 构造成和 Ψ^H 极端相似（Coherence）的矩阵，也就是拿出 Ψ 的

前M行。用这个算法求$\hat{y}$，我们将得到$\hat{y}=\begin{pmatrix} s^{M\times 1} \\ 0^{(N-M)\times 1} \end{pmatrix}$，这显然是错误的。也就是说，你强迫的认为前$M$个变换域分量是重要的。而事实是，重要的$K$个分量的位置我们事先是不知道的，是随着信号的不同而不同的。当然，你可以将Φ恰好构造成对应最重要分量的K行，得到正确的结果。而这种的做法要付出的概率代价$\frac{1}{C_N^K}$。也就是说，你必须穷举C_N^K次，才能得到你想要的结果。但是，即使你有幸碰到了它，也并不能肯定这个结果就是对的。因此，我们选择Φ和Ψ^H极端不相似（Extremely Incoherence）。于是，Φ很大程度上和随机（Randomness）这个词相联系，它可以是满足高斯分布的白噪声矩阵，或贝努里分布的± 1矩阵（也称作 Noiselet）等等。除此之外，我们希望线性测量有稳定的能量性质：$1-\delta \le \frac{\|\Phi\Psi^H \hat{y}\|_2}{\|\hat{y}\|_2} \le 1+\delta$，也就是它要保持$K$个重要分量的长度。综合上面的，我们有了如下编码解码的策略：

编码：构造Φ，生成测量$s=\Phi x$，保留s。

解码：构造同样的Φ，构造任一种正交变换Ψ^H，根据s重构x。

压缩传感的优势：1、非自适应（Non-Adaptive）的，一开始就可以传输长度较短的信号，甚至突破采样定理的极限。2、抗干扰，s中任何一项都是重要的，或者说不重要的。丢失了某几项，仍然可以完美重构。它的缺点：1、实际中，s的长度一般是重要分量长度的4倍，才能近乎完美重构。数学上更严格的，$M \approx 4K$或者$M \ge K\log_2(\frac{N}{K})$。2、重构（恢复)算法是NP问题。即使将0-范数转化为1-范数，由于其不可微性(Indifferentiable)，算法的计算复杂度仍然很高。它的应用前景广泛：低成本数码相机和音频采集设备；节电型音频和图像采集设备；天文（图像本身就稀疏，例如天空的星星)；网络；军事（用很简易

的摄像机随机记录场景，可以完全重构军事地图)；超宽带（雷达信号处理)。这里，值得指出的是，美国的工程学家已经设计出了实际的产品。

快速算法——正交匹配追踪

对于 0-范数的优化问题，实际上是 NP 问题，就是在多项式时间内难以求解，甚至无法验证解的可靠性。于是，我们必须将 0-范数换一下，变成 1-范数。为什么不是 2-范数呢，那样就会简单多了。毕竟 2-范数的优化问题可以转化成 2 次型问题，而 1-范数，$\|\hat{y}\|_1=\sum_i|\hat{y}_i|$，在 0 点处不可导，因此无论是梯度算法、矩阵求导等等手段都变得相形见绌。因此，基于 1-范数的优化算法需要特殊处理，且复杂度很高。下面我们来解释下为什么要用 1-范数，而不是 2-范数。我们令恢复（Recovery）矩阵$T=\Phi\Psi^H$，则等式约束可重写为：$T\hat{y}=s$。$\hat{y}$中未知数有N个，方程只有M个，且$M<<N$。因此，方程有无穷多解。从几何上说，$T\hat{y}-s=0$是一个超平面，为了简化，在 2-D 问题中（$K=1$，$\hat{y}$只有两个元素待求）可认为它就是一条直线。而范数约束呢？0-范数是一个十字架，因此它的最外侧（范数的最小值）是 4 个点。所以其和直线的交点，必然在坐标轴上。也就是说，能使$\hat{y}$产生更多的 0，这正是我们想要的“稀疏”的结果。2 范数是一个圆，因此它的最外侧边界和直线的交点（就是切线的概念)，以压倒性的概率不在坐标轴上，除了直线的斜率恰好为 0 或者无穷大。其实直线的斜率恰好为 0 或者无穷大，是不可能的，因为Φ和Ψ^H极端不相似。只有Φ取Ψ的某一行时，两者相似，才会发生斜率恰好为 0 或者无穷大的情况（因此，你的胜算只有$1/C_2^1$，但你不知道哪个是对的。)。依上所述，用 2-范数优化的结果，使$\hat{y}$几乎没有 0，这是我们不期望的。而 1-范数是一个菱形，四个角都在坐标轴上，因此它和直线的交点以压倒性的概率落在坐标轴上。这就是我们要用 1-范数的原因。事实上，p范数满足，$0\le p<1$，它的外边界都向中心凹（Concave)。而$p>1$，外边界向外凸（Convex)。

所以前者的外边界和直线的交点无疑落在坐标轴上。根据这个几何解释，我们可以将问题转化成：

$$\min \| s - T\hat{y} \|_2 + \lambda \| \hat{y} \|_1$$

这显然是一个非线性（Non-Linear）的凸（Convex）优化问题。

众所周知，对于优化问题，我们一般用梯度的方法来求解。而对$\| \hat{y} \|_1$，在 0 点导数不存在，因为这个点正好位于两条直线的交点上，左右导数不相等。这也正是很多数学方法的考虑。像子梯度（Subgradient）法、平滑近似法（Smooth Approximation）等等。我们暂不谈这些方法，因为它需要特殊的数学背景。我们谈一谈工程领域最常用的正交匹配追踪法（Orthogonal Matching Pursuit）。它的思想本质上还是来自于这个K“稀疏”。我们绕了一圈，还是为了找这K个关键的分量。既然是关键，显然它的系数的绝对值应该比其它$N-K$个分量大得多。为了简单起见，我们先假设$K=1$，唯一非零元素$\hat{y}_q$在$\hat{y}$中对应的位置在q。于是$T\hat{y}$就是恢复矩阵T的第q列T_q与$\hat{y}$中的非零元素$\hat{y}_q$的乘积，即$T_q \hat{y}_q = s_q$。且$\| s - s_q \|_2 / \| s \|_2 < \delta$。换句话说，$T$的第$q$列与$s$的相似程度最高，即$| < T_q, s > | = | T_q^H s | >> | T_r^H s | = | < T_r, s > |, r \neq q$。所以，我们只要计算恢复矩阵$T$的所有列与$s$的内积，找到内积绝对值最大的那列就行了，该列对应的位置就是q。根据最小二乘法，$\hat{y}_q = (T_q^H T_q)^{-1} T_q^H s$，就是使$\| s - T_q \hat{y}_q \|_2$最小的那个$\hat{y}_q$。呵呵，感觉到了吗，实际上这有点或者非常像施密特（Schimidt）正交化方法。余量$r_n = s - \frac{< T_q, s >}{< T_q, T_q >} T_q$，始终同$T_q$正交。这也是为什么这个方法叫“正交”匹配追踪的意思了。而匹配，显然是你找到了最大的$| < T_q, s > |$。对于$K>1$呢，其实是差不多的。我们找到余量r_n同T中所有列向量最大的那个即可（但第一次找到的那列要排除，因为它已经保留了下来。）。于是，找到使

$\| s-(T_{q2},T_{q1})\begin{pmatrix}\hat{y}_{q2}\\ \hat{y}_{q1}\end{pmatrix}\|_2$ 最小的那个 $\begin{pmatrix}\hat{y}_{q2}\\ \hat{y}_{q1}\end{pmatrix}$。这里，$T_{q1}$ 是我们第一次找到的那一列，T_{q2} 是我们新找到的那一列（也要记住它的列号 q_2）。可以看出，$\hat{y}_q=\begin{pmatrix}\hat{y}_{q2}\\ \hat{y}_{q1}\end{pmatrix}$ 被更新了，由原来的一个变成两个了，也就是我们找到两个在变换域最关键的元素和其在 $\hat{y}$ 中对应的位置了。令 $T_q=(T_{q2},T_{q1})$，余量 r_n 又一次被写为：$r_n=s-\dfrac{<T_q,s>}{<T_q,T_q>}T_q$。我们再一次看到了施密特正交化的影子。继续上面的步骤，直至找到变换域所有 K 个最重要的分量。也就是说，正交匹配追踪的迭代次数 $m\ge K$。实际操作上只要满足 $\| r_n\|_2/\| s\|_2<\delta$，迭代就可以中止了。最后，假设正交变换 Ψ 是傅里叶变换，我们来分析下计算复杂度。生成恢复矩阵 T，采用 FFT 快速算法，需要计算复杂度 $O(MN\log_2(N))$；找到 T 中匹配的列向量需要 $O(mNM)$ 次操作，注意 m 是算法需要的迭代次数；求解 $\hat{y}_q$ 和 r_n 需要 $O(m^2M)$ 次操作。考虑到 $m\sim K$，$M\sim K\log_2(N/K)$，总计算复杂度可以控制在 $O(N[\log_2(N)]^2)$ 以下。到此为止，我们完成了压缩传感所有基础的理论和操作。

小结——压缩传感的特点

最后，让我们总结一下，压缩传感理论的关键字：稀疏（Sparsity）、不相关（Incoherence）、随机性（Randomness）、非自适应（Non-Adaptivity）、非线性（Non-Linearity）、不可微（Indifferentiability）。从这些词语可以看出，压缩传感理论是对传统理论的颠覆。这种颠覆最令人振奋的表现，就是它突破了香农采样定理的极限，能以随机采样的方式用更少的数据采样点（平均采样间隔低于采样定理的极限），来完美地恢复原始信号。科学也就是在对传统理论不断地颠覆和修正中才得以进步和发展。

为了大家更好的理解压缩传感，在我的个人网站上，给出了一个没有经过优化的简单的操作代码，希望对大家有所帮助。此外，由于此帖含有大量公式，我用了类似 Latex 的语法

敲出了它，一个更好阅读的 pdf 版本也放在了我的个人网页上。最后，如果阅读这篇帖子的人，试图转贴的话，请注明这篇帖子的出处和作者，以此作为对作者工作小小的肯定和支持。（程序和帖子的 PDF 版下载地址：http://www.eee.hku.hk/~wsha/Freecode/freecode.htm）