文章编号:1001 - 893X(2010)11 - 0026 - 04

压缩感知测量方法的机密性*

王 超,梁大鹏

(北京科技大学 信息工程学院,北京 100083)

摘 要:分析了压缩感知(CS)的安全性问题,讨论了在攻击者不知道测量矩阵情况下是否可以有效对信号进行重构的问题,论证了压缩感知可以达到保密性但达不到完善的保密性。最后联合信道容量和速率失真函数,讨论了压缩感知恢复信号所需测量数据量的下限,并分析了测量噪声对信号重构性能的影响。

关键词:压缩感知;测量矩阵;机密性;信道容量;率失真函数

中图分类号:TP952 **文献标识码**:A doi:10.3969/j.issn.1001 - 893x.2010.11.006

Secrecy of Compressed Sensing Measurements

WANG Chao, LIANG Da-peng

(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: The security of compressed sensing (CS) is analysed. Wheather the attacker can effectively recover signal when it does not know measurement matrix is discussed. The CS can achieve the confidentiality but is fail to achieve the perfect secrecy. The lower bound of measure number for the signal recovery of CS is discussed by combining channel capacity with rate-distortion function. The influence of measurement noise on the signal reconstruction performance limitation is also analysed.

Key words: compressed sensing (CS); measurement matrix; privacy; channel capacity; rate-distortion function

1 引 言

近年来在信号处理领域中,压缩感知(Compressed Sensing,CS)作为一种新的理论备受关注。压缩感知理论框架下,处理高度可压缩的信号时,可以摒弃掉传统的采样方式而只采样对整个数据流有用的样本,然后通过解决线性规划问题来重构原始信号。压缩感知的优点在于信号的投影测量数据量远远小于传统采样方法所获的数据量,突破了香农采样定理的瓶颈,使得高分辨率信号采集成为可能。

文献[1 - 6]研究表明,利用少数稀疏线性信号的测量值来恢复稀疏信号是可行的。但是否可以利用压缩感知中的测量矩阵在对信号进行压缩采样的同时完成对信号的加密呢? 以及从信息论角度上讲,这种加密是否安全? 如果安全,在压缩采样的过程中压缩与加密同时进行,可以避免采用额外的加密方式带来的开销。这对一些特殊的应用场合是很有用的,例如在传感器网络中低功耗设备需要捕获和发送低速数据,并且传感器网络有可能布置在无人或是敌方区域,信息安全和传输的可靠性很重要。

* **收稿日期**:2010 - 07 - 15;**修回日期**:2010 - 09 - 08

基金项目:国家高技术研究发展计划(863 计划)项目(2009AA01z209);国家自然科学基金资助项目(60902042);北京市自然科学基金资助项目(4082020)

Foundation Item: The National High - Tech Research and Development Program (863 Program) of China (No. 2009AA01z209); The National Natural Science Foundation of China (No. 60902042); The Natural Science Foundation of Beijing (No. 4082020)

在有噪声的情况下,如何利用信息论框架计算出恢复数据所需的测量值的下限。本文主要讨论压缩传感的安全性问题,并分析了在噪声条件下恢复数据所需的测量数据的下限。

2　背景知识

压缩感知理论与传统奈奎斯特采样定理不同,只要信号是可压缩的或在某个变换域是稀疏的,那么就可以用一个与变换基不相关的观测矩阵将变换所得高维信号投影到一个低维空间上,然后通过求解一个优化问题就可以从这些少量的投影中高概率地重构出原信号,可以证明这样的投影包含了重构信号所需的足够信息。在该理论框架下,采样速率不决定于信号的带宽,而决定于信息在信号中的结构和内容。

最简单的模型是一个 n 维信号 X 仅含有少量的 k 个非 0 值:$X\in R^n$,$|\sup p(x)|\leqslant k<<n$,这样的信号称为 k 稀疏。一个与变换基 Ψ 不相关的观测基 Φ:$m\times n(m<<n)$对系数向量进行线性变换,并得到观测集合 Y:$m\times 1$,那么就可以利用优化求解方法从观测集合中精确或高概率地重构原始信号 X。在 CS 理论中不需要度量稀疏信号的所有 n 个值而通过在不相关基上的一小部分投影就可以恢复信号。通过计算测量向量 Y_0来度量 X(加密编码),$Y_0\in R^m$,而 Y_0是通过向量相乘 $Y_0=\Phi X$ 得到的。CS 算法的目标是近似或是精确重构 X(解密译码)。

Candes 和 Tao 指出[7]使用最小 1 范数通过解决线性方程可以重构信号:

$$(L1)\min\|x\|_1\ \text{s.t.}\ y=\Phi x \qquad (1)$$

式中,Φ 满足约束等距性(RIP)。

目前为止,出现的重构算法都可归入以下三大类[8]:

(1)贪婪追踪算法:这类方法是通过每次迭代时选择一个局部最优解来逐步逼近原始信号,这类算法包括 MP 算法、OMP 算法、正则化 OMP(ROMP)算法、CoSaMP 算法、SP 算法;

(2)凸松弛法:这类方法通过将非凸问题转化为凸问题求解找到信号的逼近,如 BP 算法、内点法和迭代阈值法;

(3)组合算法:这类方法要求信号的采样支持通过分组测试快速重建,如傅里叶采样、链式追踪等。

3　无噪声情况下 CS 机密性

为了讨论压缩感知测量的机密性能,引入以下模型。对于一个 k 稀疏的信号 $x\in R^n$,密钥 $i\in\{1,2,3,\cdots,S\}$对应于 $m\times n$ 矩阵 $\boldsymbol{\Phi}_i$。在这个模型中,Alice 想要向 Bob 发送一则加密的信息。Alice 选择 i 使用 $\boldsymbol{\Phi}_i$和 $y=\boldsymbol{\Phi}_i x$ 来加密 x。只有密文 y 传送给 Bob,Bob 方知道加密用的密钥。给定 $\boldsymbol{\Phi}_i$、y 以及 x 的稀疏度和密钥知识,Bob 可以重构 x。Eve 窃听到 y,但他不知道加密用的 Key,也就是 i。接下来我们讨论在 Eve 只知道 y、信号 x 的稀疏度、密钥组和相应的 $\boldsymbol{\Phi}$ 矩阵来重构 x 的难度。

“一次一密”系统在理论上被认为是不可破译的,而压缩传感系统可以作为一次一密系统,实际的加密系统是通过双方共享重复使用的有限长度的主密钥,利用算法复杂性产生伪随机数进行加密。

在实际中,通过共享一个足够大的随机种子产生器,将会使密钥数量 S 达到足够大,而编码加密所用密钥随机从密钥数 S 中选取,这可以当作一次一密,而且由于对所有密钥逐一估算很困难,可以认为此系统具有安全机密性。文献[9]中指出由于 $y=\boldsymbol{\Phi}x$,即 y 和 x 存在相关性,因此有 $P_{X|Y}(X|Y)\neq P_X(X)$,则压缩感知达不到完善保密性。

随机产生的 $m\times n$ 高斯随机矩阵$\boldsymbol{\Phi}$ 和$\boldsymbol{\Phi}'$,对于 k 稀疏向量 x 满足 $y=\boldsymbol{\Phi}x$,当 $m>k+1$ 时,基于 $\boldsymbol{\Phi}$ 和$\boldsymbol{\Phi}'$所有满足 $y=\boldsymbol{\Phi}x'$的 x'是 m 稀疏。

首先对于 $\boldsymbol{\Phi}$ 和$\boldsymbol{\Phi}'$的 m 列组有唯一的恢复结果,$\boldsymbol{\Phi}'$ 的 m 列标记为 $\boldsymbol{\Phi}_m$,$\boldsymbol{\Phi}_m$与 $\boldsymbol{\Phi}'$ 相互独立,则 $\boldsymbol{\Phi}'_{\Omega m}$的秩为 m,并且此矩阵的逆矩阵可以唯一决定 x 的 m 个值,使 $y=\boldsymbol{\Phi}'x$。

最后说明当 $t<m$ 时,测量的 t 列不足以精确恢复信号。Ω_t 代表 t 稀疏向量 θ' 的非零值,则 $\boldsymbol{\Phi}_{\Omega t}$在$\boldsymbol{\Phi}'$上秩为 t。Ω_k 代表 k 稀疏向量 θ 的非零值,则 $\boldsymbol{\Phi}_{\Omega k}$在$\boldsymbol{\Phi}$ 上秩为 k。colspan($\boldsymbol{A}$)指向量 $\boldsymbol{A}$ 的列向量跨越空间,当 $m>k+1$ 时,使用聚合矩阵[$\boldsymbol{\Phi}_{\Omega k}$,$\boldsymbol{\Phi}_{\Omega t}$];又当 $k+t>m$ 时,聚合矩阵在 $\boldsymbol{\Phi}$ 和$\boldsymbol{\Phi}'$秩为 m,交集部分为 $k+t-m$。由于 $t<m$ 则交集部分维数小于k,测量向量如果在交集部分则只能用 $\boldsymbol{\Phi}_{\Omega t}$表示,但实际上测量向量存在于 k 维空间,所以 t 稀疏向量 x'在$\boldsymbol{\Phi}'$上满足 $y=\boldsymbol{\Phi}'x'$的概率为 0。当 $k+t<m$ 时聚合矩阵秩为 $k+t$,所以交集部分为 0,因此有 $y\notin$ colspan($\boldsymbol{\Phi}_{\Omega t}$),则对于 k 稀疏的信号,当 $k>m+1$ 时,使用错误密钥进行信号重构得到的将

是 m 稀疏信号而不是 k 稀疏信号。

Eve 对 k 稀疏信号使用错误的密钥进行重构得到的信号将是 m 稀疏的,因此利用 1 范数时可以认为该压缩感知测量具有保密性。

4 信号重构性能限制

率失真函数理论指出为了达到在目标数据率的条件下使传输信号的失真最小,在编码比特率和信号失真之间必须选择一个恰当的折衷,这是香农信息论中的率失真理论问题[10]。率失真理论讨论的主要问题是:在允许一定程度失真的条件下,能够把信息压缩到什么程度。若定义最大允许失真度为 D,则其对应的编码比特率的下限 $R(D)$是 D 的单调递减函数,称为率失真函数。率失真函数的定义表明:一个具有率失真函数 $R(D)$的信源,倘若 $R<R(D)$,则不存在任何编码会使失真度小于 D。或者说,没有任何一种编码方法可以使平均编码比特率小于 $R(D)$,而又使失真度小于 D。所以,率失真函数提供了一把衡量实际所采用的编码方法效率是高或低的尺子。由此可见,率失真理论给我们提供了达到最佳效果的编码原则。

对于压缩感知理论,假设 X 是时间离散幅度连续的信源,测量值 Y 可以看作高斯信道的输出,无噪情况下的 Y_0作为信道输入。由于信道容量有限则每个测量仅获得有限的信息量,也说明了完美信号重构是不可能的。由率失真函数理论可知,使用 CS 测量重构方案重构信号 X,并要求达到一定的保真度,则最小测量速率需要达到一定大小。

4.1 求信道容量 C

CS 测量值所获得信息是由测量信道的容量决定的,要寻找特定失真率下测量速率 $\delta(\delta=m/n,n$ 为信号长度)的下限,就要先求得信道容量。

对于 $Y=Y_0+Z=\Phi X+Z$,Y_0是理想(无噪)状况下的测量向量,Z 是属于 R^m 的包含 m 个 0 均值、1 方差($\sigma=1$)的高斯随机变量。作如下定义:

信噪比:

$$SNR=E[\|Y_0\|_2^2]/E[\|Z\|_2^2]=E[\|Y_0\|_2^2]/m \tag{2}$$

失真率:

$$D=E[\|X-X'\|_2^2]/E[\|X\|_2^2],\delta=m/n \tag{3}$$

文献[11]给出了信道容量的公式:

$$C=\max_{tr(\sum Y_0)\leqslant mSNR}\frac{1}{2m}\lg\frac{|\sum Y_0+\sum Z|}{|\sum Z|} \tag{4}$$

式中,$\sum Y_0$、$\sum Z=\boldsymbol{I}_{m\times m}$分别指代 Y_0和 Z 的协方差矩阵,信噪比 $SNR=(1/m)\,\mathrm{tr}(\sum Y_0)$,其中$\mathrm{tr}(\cdot)$是对矩阵的遍历,$C$ 是以比特计算的每个测量的信道容量。

又有$\sum Z=\boldsymbol{I}_{m\times m}$,所以$|\sum Z|=1$,则有:

$$C=\max_{tr(\sum Y_0)\leqslant mSNR}\frac{1}{2m}\lg(|\sum Y_0+\boldsymbol{I}_{m\times m}|) \tag{5}$$

为了最大化信道容量使用 Hadamard 不等式,正定矩阵的行列式小于等于其对角元素的乘积[12],即$|\wedge|\leqslant\prod_i\wedge(i,i)$,所以有:

$$|\sum Y_0+I_{m\times m}|\leqslant\prod_i(1+\sum Y_0(i,i)) \tag{6}$$

继而有:

$$C\leqslant\max_{tr(\sum Y_0)\leqslant mSNR}\frac{1}{2m}\lg(\prod_i(1+\sum Y_0(i,i)))\leqslant\frac{1}{2}\lg(1+SNR) \tag{7}$$

当 Y_0是对角矩阵且对角元素都等于 SNR 时上式取等号,所以最好的 CS 测量系统是测量向量相互独立并且方差相同。通过信道容量 C 可以从 m 个有噪测量值 Y 中获取最大信息。

4.2 使用信源信道分离定理来估计误差界

根据文献[12],对于时间离散幅度连续平稳遍历信号,信源 X 经 m 个信道传播,当且仅当从信道获取的信息量 mC 大于量化信源信息量$nR(D)$时 X 的失真度为 D。

4.3 计算结果

通过上面的分析可知,达到 D 误差率的 CS 测量速率的下限为

$$\delta\geqslant 2R(D)/\lg(1+SNR) \tag{8}$$

式中,$R(\cdot)$是速率失真函数。对于高斯信源速率失真函数为

$$R(D)=\frac{1}{2}\lg(\sigma^2/D) \tag{9}$$

所以由以上两式得到 $\delta\geqslant\lg(\sigma^2/D)/\lg(1+SNR)$即为恢复信号所需测量值数量的下限,同样可以看出测量速率对于测量信噪比 SNR 有重要的影响。

利用信息论理论讨论在噪声环境下 CS 恢复信号所需测量数量的下限。主要思想是将噪声情况下信号获取过程当作通信信道模型处理,信道容量表示出测量值所包含的信息量,使用这个结果和信源率失真函数可以得到所需的测量速率。

5 总　结

本文在压缩感知理论基础上,讨论了无噪情况下采集压缩感知数据的安全性问题,指出其具有保密性但达不到完善保密性。利用信息论知识给出了有噪的情况下,CS 恢复信号所需测量数量的下限。有噪情况下 CS 的安全性能问题将是进一步研究的内容。

参考文献:

[1] 石光明,刘丹华,高大化,等.压缩感知理论及其研究进展[J].电子学报,2009,37(5):1071 – 1081.
SHI Guang – ming, LIU Dan – hua, GAO Da – hua, et al. Advances in theory and application of compressed sensing[J]. Acta Electronica Sinica, 2009, 37(5): 1071 – 1081. (in Chinese)

[2] Candes E, Romberg J, Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information [J]. IEEE Transactions on Information Theory, 2006, 52(2): 489 – 509.

[3] Candes E, Tao T. Near optimal signal recovery from random projections: Universal encoding strategies? [J]. IEEE Transactions on Information Theory, 2006, 52(12): 5406 – 5425.

[4] Donoho D. Compressed Sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289 – 1306.

[5] Baraniuk R G. Compressive sensing[J]. IEEE Signal Processing Magazine, 2007, 24(4): 118 – 121.

[6] Candes E, Wakin M. An introduction to compressive sampling [J]. IEEE Signal Processing Magazine, 2008, 25(2): 21 – 30.

[7] Candes E J, Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory, 2005, 51(12): 4203 – 4215.

[8] Needell D, Tropp J A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples [J]. Comp Harmonic Anal, 2009, 26(3): 301 – 321.

[9] Rachlin Y, Baron D. The secrecy of compressed sensing measurements[C]//Proceedings of Allerton Conference on Communication Control and Computing. Urbana-Champaign, IL: IEEE, 2008: 813 – 817.

[10] 傅祖芸.信息论——基础理论与应用[M].北京:电子工业出版社,2001.
FU Zu – yun. Information theory——basic theory and application [M]. Beijing: Publishing House of Electronics Industry, 2001. (in Chinese)

[11] Cover T M, Thomas J A. Elements of Information Theory [M]. New York: Wiley Press, 1991.

[12] Berger T. Rate Distortion Theory: A Mathematical Basis for Data Compression[M]. Englewood, NJ: Prentice – Hall, 1971.

作者简介:

王　超(1978 –),男,河北人,博士,北京科技大学信息工程学院讲师,主要研究方向为无线通信、信号处理;

WANG Chao was born in Hebei Province, in 1978. He is now a lecturer with the Ph. D. degree. His research interests include the wireless communication networks and signal processing.

Email: wanch3307@sohu.com

梁大鹏,硕士研究生,主要研究方向为无线通信网络和认知无线电。

LIANG Da – peng is now a graduate student. His research interests include the wireless communication networks and cognitive radio.