

# Sparsity Estimation Method Based on Noise Characteristic Analysis

Yimin Jiao, Chunyu Kang, Xiangxu Zeng, Tianyu Liu

Dalian Naval Academy, Dalian Liaoning  
Email: 412293775@qq.com

Received: Apr. 3<sup>rd</sup>, 2016; accepted: Apr. 17<sup>th</sup>, 2016; published: Apr. 21<sup>st</sup>, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The sparse representation, after decades of development, has been deeply studied and applied in many fields. Estimation of signal sparsity is an important part of the work in the sparse decomposition of signal. According to the noise characteristics, this paper proposes a method for estimating sparse degree. By constructing the Fourier base dictionary, using the uniform distribution characteristics of noise energy in the whole frequency domain, and through the sparse characteristics of signal traversing the full spectrum, the accuracy of the signal sparsity is gradually determined. Simulation results show that the proposed method can effectively complete the estimation of signal sparsity.

## Keywords

Sparse Representation, Sparsity Estimation, Orthogonal Matching Pursuit Algorithm

---

# 基于噪声特性分析的稀疏度估计方法

焦义民, 康春玉, 曾祥旭, 刘天宇

海军大连舰艇学院, 辽宁 大连  
Email: 412293775@qq.com

收稿日期: 2016年4月3日; 录用日期: 2016年4月17日; 发布日期: 2016年4月21日

## 摘 要

稀疏表示方法经过几十年的发展, 在很多领域得到了深入研究与应用。信号稀疏度的估计是对信号进行稀疏分解的工作中的重要环节。根据噪声的特性, 提出了一种稀疏度估计方法。即通过构造傅立叶基字典, 利用噪声能量在全频域均匀分布的特点, 通过遍历信号在全频谱上的稀疏特性, 逐步确定准确的信号稀疏度。仿真实验结果表明, 提出的方法能有效的完成对信号稀疏度的估计。

## 关键词

稀疏表示, 稀疏度估计, OMP算法

## 1. 引言

对信号稀疏度的估计一直是稀疏表示领域中需要迫切解决的问题, 当我们知道了信号稀疏度时, 可以更为准确的对信号进行稀疏表示。目前对信号的稀疏度估计的方法有主成分变换(Principle Component Analysis, PCA)方法[1]-[3]、自适应稀疏度估计的贪婪回溯(Adaptive Sparsity Matching Pursuit, ASMP)方法[2] [4] [5]等是目前比较常用的方法。

前者通过类似小波变换方法估计图像稀疏度的方法设计出利用主成分分析变换来估计图像稀疏度, 主成分变换方法表示具有速度快、复杂度低等优点, 因此利用主成份分析方法进行稀疏度估计的结果评价更客观。文章假设图像的稀疏度与其方差存在线性关系, 且该线性关系可以通过数据拟合得到, 在仿真中利用主成份分析方法计算图像的方差, 利用拟合得到的线性关系估计图像的稀疏度, 仿真测试结果显示这种方法取得了很好的稀疏度估计效果。但该方法依然存在不足, 一是没有充分的理论依据对文中的假设做出说明, 二是仿真测试选择的图像数量过少。

后者是通过采取一种新的稀疏表示匹配算法 ASMP 算法, 迎合自然条件下稀疏度未知的条件, 弥补文中算法的不足, 该算法不仅互补了 MP, OMP 等匹配算法中的缺陷, 在算法复杂度及鲁棒性方面都有所增强。同时在图像的稀疏分解中, 综合 StOMP 算法和 CoSaMP 算法的优势, 互补了以上两种算法的缺陷, 在提高算法鲁棒性和可靠性的同时折中了传统稀疏分解去噪算法的时间开销。

以上两种方法主要应用于图像处理领域。但在一维信号处理领域, 对信号稀疏度的估计的要求和图像压缩不同。本文尝试通过探求信号中的噪声特性, 根据噪声的分布特性, 在遍历信号在全频域的的稀疏特性的基础上, 构造一种估计信号稀疏度的方法。利用信号在冗余稀疏度的稀疏分解下的相对误差的分布, 逐步准确的估计信号在噪声下的稀疏度。并通过仿真数据对提出的方法进行验证。

## 2. 稀疏基与稀疏表示算法

### 2.1. 傅立叶基字典的构造

傅立叶变换能够得到信号的频域表示, 反映了信号在全部时间范围内的所有频谱成分, 因此, 傅立叶变换在描述平稳信号时效果很好。如果将傅立叶变换以基的形式描述, 则傅立叶基是频域字典中最典型的一个完备字典, 这个字典的原子可用正弦波性的集合  $g_\gamma$  表示, 如式(1)所示,  $f$  表示原子的频率,  $nT_s$  表示信号时间[6] [7]。

$$g_\gamma(n) = e^{2\pi jfnT_s} \quad (1)$$

其中  $n \in \{0, 1, \dots, N-1\}$ 。

## 2.2. 正交匹配追踪算法

正交匹配追踪算法在稀疏分解中的应用最广泛，它的原理是在一般匹配追踪的基础上增加正交化过程，这个过程可以用更少的原子来表示信号，使信号变得更为稀疏，因此 OMP 算法的收敛速度较快[8]。

OMP 算法用于信号的稀疏表示的具体步骤如下：

**步骤 1:** 设定算法输入，信号  $y$ ，字典  $A$ ，稀疏度  $K$ 。

**步骤 2:** 初始化算法参数，残差  $r_0 = y$ ，分解系数  $\hat{x} = 0$ ，索引集  $t_0 = \emptyset$ ，子字典  $T_0 = \emptyset$ ，迭代因子  $l = 1$ ，最大迭代次数  $\text{iterNum} = K$ 。

**步骤 3:** 迭代过程，在第  $l$  次循环 ( $l \geq 1$ )，运行以下分步骤。

1) 相关最大计算寻找最佳原子索引

$$i_l = \arg \max_i \left( \left| \langle d_i, r_{l-1} \rangle \right| \right), i = 1, 2, \dots, N$$

2) 更新原子索引集合

$$t_l = t_{l-1} \cup i_l$$

3) 更新子字典

$$T_l = T_{l-1} \cup d_{i_l}$$

4) 更新系数估计

$$\hat{x}[t_l] = T_l^+ y$$

其中， $T_l^+ = (T_l^H T_l)^{-1} T_l^H$ 。

5) 更新残差

$$r_l = y - T_l (T_l^+ y)$$

6) 判断终止条件，若  $l > K$ ，算法结束。

**步骤 4:** 输出分解系数  $\hat{x}$ 。

**步骤 5:** 合成稀疏信号

$$y = A \hat{x}$$

## 3. 稀疏度估计方法

### 3.1. 基于主成份分析的稀疏度估计方法[1] [3]

对信号进行 PCA 变换后的系数特性进行分析，假设对于任意给定的自然场景图像，其 PCA 后的系数为  $\alpha_i, i = 0, 2, \dots, n-1$ ，则该系数曲线可近似看作正态分布函数，其均值为  $\mu$ ，方差  $\sigma$  可通过下式得到。

$$\sigma = \sqrt{\frac{\sum_{i=0}^{n-1} (\alpha_i - \mu)^2}{n}}$$

稀疏度  $K$  与方差  $\sigma$  间存在如下的线性关系

$$K = a\sigma + b$$

式中， $a$  和  $b$  为常数，可以通过数据拟合的方式得到可以看出，当上述假设成立时，只需要对信号数据

进行一次 PCA 得到变换后的系数，就可以通过上述两式对稀疏度进行估计。

将得到的系数采用最小二乘法进行拟合，拟合时直接采用 Matlab7.x 中的 polyfit 函数，该函数采用最小二乘法曲线拟合方法，所得到的函数值在基点处的值与原来点的坐标偏差最小。得到拟合公式为

$$K = 0.3607 \times \sigma + 2.5499 \quad (2)$$

### 3.2. 基于噪声特性的稀疏度估计方法

在无噪声干扰的情况下，OMP 算法会将信号  $S$  在字典  $A$  的作用下稀疏表示为指定稀疏度  $K$  的稀疏信号  $Y$ 。当指定稀疏度  $K$  小于信号  $S$  的真实稀疏度时，OMP 算法会将信号  $S$  在频域中按频率分量的大小，从大到小选出  $K$  组频率分量输出为稀疏信号  $Y$ ；当指定稀疏度  $K$  大于或等于信号  $S$  的真实稀疏度时，理论上稀疏信号  $Y$  和信号  $S^f$  相等。在有噪声干扰的情况下，和无噪声干扰不同的是当指定稀疏度  $K$  大于或等于信号  $S$  的真实稀疏度时，稀疏信号  $Y$  和信号  $S$  不再相等，而是随着指定稀疏度  $K$  的增大，稀疏信号  $Y$  不仅会信号  $S^f$  中的所有分量，而且还会有“杂波”。这些杂波本质上是噪声，在 OMP 算法中的过稀疏度表示时被误认为是信号有用频率分量。而且这些杂波的出现也有规律，即杂波出现的频率分量刚好是白噪声在该处能量最大的地方。噪声在整个频谱近似于均匀分布，即杂波出现的频点位置以均匀分布的方式依概率分布在全频域中。这个性质为我们估计稀疏度提供了理论支撑[9] [10]。

于是我们想到，可否现将信号  $S$  进行分段，然后每一段信号进行一次稀疏度为  $K$  的 OMP 分解，根据分解的结果，来判断稀疏度选择的正确性。若所有分段后的信号  $S^f$  都相等，说明稀疏度选择过小，若不等，说明稀疏度选择过大，重复多次直到确定信号  $S$  的稀疏度。

但单单依靠这个方法去估计稀疏度会存在一个问题，在逐步增大稀疏度  $K$  时，有可能在噪声较大时，估计的稀疏度偏小。这是因为 OMP 算法在对信号进行稀疏表示时，每次迭代只选取一个能量最大的频率分量造成的。当信号  $S^f$  中有  $N$  个频率分量上的能量相差无几时，在噪声的干扰下，会随机的使其中一个频率分量的能量超过其他的频率分量，导致 OMP 算法在分解后出现不相同的  $S^f$ ，但这时指定稀疏度并未达到信号的真实稀疏度，而导致估计错误，我们称这类估计错误的原因第二类误差。

针对这个问题，需要寻求一种新的稀疏度判别手段，对分段后的信号  $S^f$  的频点位置进行统计，若新出现的频率分量的位置个数小于信号分段数的  $\frac{1}{N}$ ，我们就有很大的把握相信出现了上面所说的第二类误差。若新出现的频率分量的位置数大于信号分段数的  $\frac{1}{N}$ ，这时我们有很大的把握相信预定的稀疏度选择过大。在分段数的选择上，一般选择大于信号  $S^f$  中最大  $N$  的三倍。

通过前面分析，信号稀疏度估计的步骤如下：

**步骤 1：** 将信号“ $S$ ”进行分段为  $S_1, S_2, \dots, S_{num}$ 。

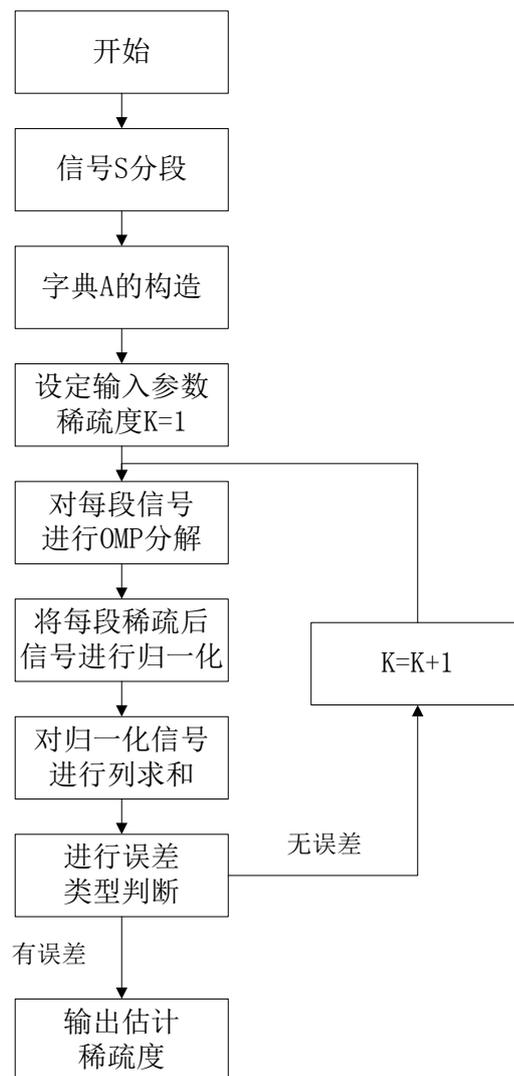
**步骤 2：** 构造傅立叶基字典  $A$ 。

**步骤 3：** 假定稀疏信号  $S_1, S_2, \dots, S_{num}$  的稀疏度  $K$  为 1。对  $S_1, S_2, \dots, S_{num}$  依次进行 OMP 算法稀疏表示为  $Y_1, Y_2, \dots, Y_{num}$ 。即将  $Y_1, Y_2, \dots, Y_{num}$  组成一个矩阵  $Y_{sp} = [Y_1, Y_2, \dots, Y_{num}]^T$ ，并将  $S_{sp}$  中非零元素全部变为 1，组成矩阵  $Y_{sp}^*$ ，然后对矩阵  $Y_{sp}^*$  列求和得到  $Y_{sumsp}^* = [Y_{col1}, Y_{col2}, \dots, Y_{colnum}]^T$ ，计算  $Y_{sumsp}^*$ 。满足条件  $0 < sum_i < 0.9, i = 1, 2, \dots, num$  的元素个数  $M$ ，若存在大于等于分段数的个数  $M$ ，则判断有误差，否则判断无误差。

**步骤 4：** 若存在误差，则执行步骤 5，否则将稀疏度加 1，再一次对  $S_1, S_2, \dots, S_{num}$  依次进行稀疏表示，执行步骤 3。

**步骤 5：** 将  $K-1$  作为估计后的稀疏度输出。

实现框图如下：



## 4. 实验仿真验证

### 4.1. 基于主成份分析法的稀疏度估计方法验证

假设仿真信号由单频信号 120 Hz 起，步长为 4 Hz，终止频率为 132 Hz 到 154 Hz 组成的 6 组正弦信号，运用上述(2)式进行稀疏度估计。估计结果如表 1 所示。

表 1 显示了在一维信号中，利用主成份分析方法对稀疏度进行估计的结果，结果显示该方法并不能有效的对一维信号进行有效的稀疏度估计。

### 4.2. 基于噪声特性分析法的稀疏度估计准确率随信噪比的变化

假设仿真信号由单频信号 120 Hz 起，步长为 4 Hz，终止频率为 160 Hz 组成的正弦信号，幅值为 5, 6, 1, 6, 4, 1, 2, 4, 6, 6; 经过高斯白噪声加扰形成。采样频率  $F_s = 2500$  Hz，字典 A 起始频率为 0 Hz，终止频率为 250 Hz。误差门限设置为 1 倍信号分段数。信号长度为 25,000，将该信号均分为 10 段，进行 20 次稀疏度估计。

运用本文的稀疏度估计方法对该仿真信号进行稀疏度估计准确率随信噪比变化如图 1 所示。

Table 1. System resulting data of sparsity estimation

表 1. 稀疏度估计结果数据

信号真实稀疏度	$\sigma$	估计稀疏度	相对偏差
1	7.1777	5.1389	413.89%
2	8.6330	5.6638	183.19%
3	6.3652	4.8458	61.52%
4	10.0767	6.1846	54.61%
5	8.8899	5.7565	15.13%
6	8.3682	5.5683	7.19%

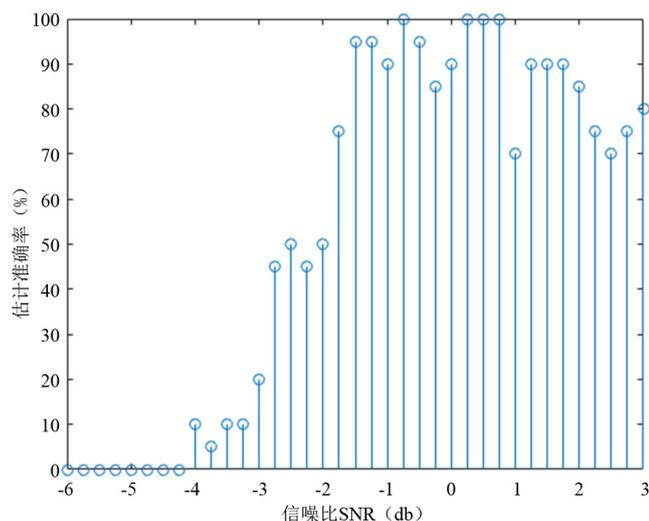


Figure 1. Stem: system result of sparsity estimation

图 1. 标准试验系统结果线图

从图 1 可以看出, 当信噪比小于  $-4$  db 时, 信号稀疏度估计的准确率为零, 此时信号受噪声干扰较为严重。当信噪比在  $-4$  db 至  $-1.5$  db 时, 信号稀疏度估计的准确率随着信噪比的增大而增大。当信噪比在  $-1.5$  db 至  $0.75$  db 时, 信号稀疏度估计的效果最好, 可以达到 85% 以上。当信噪比大于  $0.75$  db 时, 稀疏度估计的准确率不升反降, 这是因为当信噪比过高时, 起始频率与终止频率之差是 250 Hz, 在稀疏度为 10 的仿真信号中, 误差稀疏度会以一定的概率重复出现在同一频点, 导致估计稀疏度比真实稀疏度高 1, 导致稀疏度估计的准确率在 90% 左右。

## 5. 结论

本文利用含噪信号在全频域上的稀疏特性, 构造了基于傅立叶基的完备原字库。再根据噪声特性, 在冗余稀疏度条件下对含噪信号进行稀疏表示, 运用本文的误差判别方法对信号稀疏度进行估计。实验表明, 本方法可以在较低的信噪比噪声环境下对信号的稀疏度进行估计。相比之下, 本文方法比利用主成分分析方法对信号稀疏度估计更为准确, 在低信噪比的情况下依然可以对信号的稀疏度进行有效的估计。

## 基金项目

国家自然科学基金资助(61471378)。

## 参考文献 (References)

- [1] 马原, 吕群波, 刘扬阳, 钱路路, 裴琳琳. 基于主成分变换的图像稀疏度估计方法[J]. 物理学报, 2013, 62(20): 204-202.
- [2] Kim, E. and Paul, G. (1987) *Principal Component Analysis*. Elsevier Science Publishers, Amsterdam, 37-52.
- [3] Meyer, Y. (1993) *Wavelets and Operators*. Cambridge University Press, Cambridge.
- [4] 张鹏飞, 金晓康, 汤倩, 张建明. 学习字典下自适应稀疏度估计的分解去噪算法[J]. 福建电脑, 2015(1): 6-8.
- [5] 徐勇俊. 基于信号稀疏表示的字典设计[D]: [硕士学位论文]. 南京: 南京理工大学, 2013.
- [6] 郭俊锋, 郑晓慧, 魏兴春. 滚动轴承振动信号的稀疏表示研究[J]. 甘肃科学学报, 2014, 26(3): 91-94.
- [7] 彭志珍. 基于改进的匹配追踪算法的信号稀疏分解[D]: [硕士学位论文]. 武汉: 华中科技大学, 2011.
- [8] 沈益青. 匹配追踪算法中稀疏度的自适应研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2013.
- [9] Traub, J.F. and Wozniakowski, H. (1980) *A General Theory of Optimal Algorithms*. Academic, New York.
- [10] Donoho, D.L., Elad, M. and Temlyakov, V. (2006) Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. *IEEE Transactions on Information Theory*, **52**, 6-18.