自由分布式 FDR 假设检验阈值算法的小波域信号增强算法

张天瑜
(无锡市广播电视大学机电工程系，江苏无锡 214011)

摘要：在小波阈值增强算法中存在的主要问题是阈值的选取。Donoho 提出的通用阈值在实际的应用中效果并不十分理想。利用统计学中的自由分布式错误发现率(False Discovery Rate, FDR)假设检验方法，提出一种新的阈值算法。选取 sym4 小波函数对带噪信号进行处理，仿真结果表明该算法的信号增强效果优于通用阈值的信号增强效果。

关键词：小波去噪；信号增强；通用阈值；错误发现率；小波阈值
中图分类号：TN912.35 **文献标识码**：A **文章编号**：1000-3630(2010)-01-0078-05
DOI 编码：10.3969/j.issn1000-3630.2010.01.018

Research on signal enhancement algorithm in wavelet domain with threshold algorithm of free distributed FDR hypotheses test

ZHANG Tian-yu
(*Department of Mechanical and Electrical Engineering, Wuxi Radio & Television University, Wuxi* 214011, *Jiangsu*, *China*)

Abstract: The main problem of wavelet threshold enhancement algorithm is threshold selection. The universal threshold proposed by Donoho is not very ideal in practice. According to the method of free distributed FDR hypotheses testing in statistics, a new threshold algorithm is proposed. Sym4 wavelet function is selected to deal with the noise signal. The simulation shows that the proposed algorithm has better effects than that of the universal threshold on signal enhancement.
Key words: wavelet denoising; signal enhancement; universal threshold; FDR; wavelet threshold

1 引 言

小波阈值增强算法是小波去噪领域中的一个热点。目前国外有很多种确定阈值的方法，其中主要有通用阈值法(Sqtwolog 规则)、Stein 无偏风险阈值法(Rigrsure 规则)、启发式 Stein 无偏风险阈值法(Heursure 规则)、极大极小阈值法(Minimaxi 规则)[1-6]。虽然 Donoho 在理论上证明并找到了最优的通用阈值，但在实际的应用中效果并不十分理想。

本文利用统计学中的自由分布式 FDR 假设检验方法，提出一种新的阈值算法，即自由分布式 FDR 假设检验阈值算法，简称 FDR 算法。

2 小波阈值增强算法

2.1 小波阈值增强算法

小波阈值增强算法可以简单地描述为三步：首先对带噪信号进行小波变换；其次对变换后的小波系数进行阈值化处理；最后对处理后的小波系数进行小波逆变换，重构原始信号。其中第二步是小波阈值增强算法的核心。

小波阈值增强算法的数学模型为：

$$f(k)=s(k)+n(k),\ k=0,1,2,3,\cdots,N-1 \tag{1}$$

式中，$s(k)$ 为原始信号，$n(k)$ 为高斯白噪声，$f(k)$ 为带噪信号。

对带噪信号进行小波变换，可得：

$$W(f)=W(s)+W(n) \tag{2}$$

因此，小波阈值增强算法三个步骤可描述为：

$$w=W(f),\ w_\lambda=D(w,\lambda),\ \hat{g}=W^{-1}(w_\lambda) \tag{3}$$

式中，$W(*)$ 为小波变换算子，$D(*,*)$ 为对带噪信号的小波系数进行阈值化处理的算子，主要包括阈值的设定和阈值函数的选取，$W^{-1}(*)$ 为小波逆变换算子。

2.2 小波阈值增强算法中通用阈值的设定

对带噪信号进行增强，特别是当噪声为高斯白噪声时，Donoho 提出了两个增强的前提条件[7]：

收稿日期：2008-10-31；**修回日期**：2009-01-06
基金项目：无锡市广播电视大学校级课题基金资助
作者简介：张天瑜(1980-)，男，江苏无锡人，硕士，讲师，研究方向为小波理论及信号去噪。
通讯作者：张天瑜，E-mail: zhangtianyu163@tom.com

(1) 光滑性：在大概率情况下，增强后的信号 $\hat{f}$ 至少和原始信号 f 有同样的光滑度；

(2) 适应性：增强后的信号 $\hat{f}$ 是最小均方差估计。

由条件(1)可以推出，当信号长度 $N\to\infty$ 时，式(4)以接近于 1 的概率成立：

$$\|\hat{f}\|_F \le C_1\|f\|_F \tag{4}$$

式中，C_1 为一大于零的常数。由此可得：

$$|\hat{\theta}_{jk}| \le |\theta_{jk}| \tag{5}$$

式中，k 为位置，$1\le k\le N$，j 为尺度，θ_{jk} 和 $\hat{\theta}_{jk}$ 分别表示原始信号和增强后的信号在尺度 j 上的第 k 个小波系数。

条件(2)可理解为对 $E(\frac{1}{N}\sum_{k=1}^{N}|\hat{f}_k-f_k|^2)$ 求最小值，等价于求 $E\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_F$ 的最小值。Donoho 证明，此时 $\hat{f}$、$\hat{\boldsymbol{\theta}}$ 必须满足：

$$E\|\hat{s}-s\|_F^2 \le \gamma E\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_F^2 \le \gamma\sigma N\sqrt{2\ln N} \tag{6}$$

当小波变换为正交小波变换时，γ=1，有：

$$\frac{1}{N}E\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_F^2 \le \sigma\sqrt{2\ln N} \tag{7}$$

由式(7)可知，对于任意 $|\theta_{jk}|\le\sigma\sqrt{2\ln N}$，取 $\hat{\theta}_{jk}$=0，将满足式(6)。此时，θ_{jk} 由噪声产生，因此，通用阈值可设为：

$$\lambda=\sigma\sqrt{2\ln N} \tag{8}$$

由上述推导可看出，该小波系数中最大值小于 λ 的概率趋近于 1，即存在一个通用阈值 λ，使得该序列中的所有小波系数都小于它。随着小波分解尺度的逐渐增大，小波系数的长度将变得越来越短，从而导致由式(8)所决定的通用阈值也变得越来越小。因此，当原始信号中带有高斯白噪声时，可通过设置通用阈值来去除噪声，以达到信号增强的效果。

2.3 小波阈值增强算法中阈值函数的选取

在小波阈值增强算法中，阈值函数的选取分两种形式，分别是软阈值函数和硬阈值函数。

(1) 软阈值函数：

$$\hat{\tilde{\theta}}_{jk}=\begin{cases}\tilde{\theta}_{jk}-\lambda & \tilde{\theta}_{jk}\ge\lambda\\ 0 & |\tilde{\theta}_{jk}|<\lambda\\ \tilde{\theta}_{jk}+\lambda & \tilde{\theta}_{jk}\le-\lambda\end{cases} \tag{9}$$

将带噪信号的小波系数 $\tilde{\theta}_{jk}$ 与所选定的阈值 λ 进行比较，大于等于阈值的点取该点与阈值的差；小于等于阈值相反数的点取该点与阈值的和；幅值小于阈值的点取为零。

(2) 硬阈值函数：

$$\hat{\tilde{\theta}}_{jk}=\begin{cases}\tilde{\theta}_{jk} & |\tilde{\theta}_{jk}|\ge\lambda\\ 0 & |\tilde{\theta}_{jk}|<\lambda\end{cases} \tag{10}$$

将带噪信号小波系数的绝对值 $|\tilde{\theta}_{jk}|$ 与所选定的阈值 λ 进行比较，小于阈值的点取为零；大于等于阈值的点保持不变。

通过式(9)和式(10)的对比可以看出，用软阈值函数得到的小波系数 $\hat{\tilde{\theta}}_{jk}$ 整体的连续性较好，从而使重构的信号不会产生附加震荡，但是当 $|\tilde{\theta}_{jk}|\ge\lambda$ 时，$\hat{\tilde{\theta}}_{jk}$ 和 $\tilde{\theta}_{jk}$ 总是存在恒定的偏差，这将直接影响重构信号与原始信号的逼近程度。硬阈值函数在均方误差意义上优于软阈值函数，但是由于其在 λ 处不连续，从而导致重构信号会产生附加震荡，不具有同原始信号一样的光滑性。总体看来，软阈值函数会使去噪后的信号略显平滑，但是会丢失信号的某些特征；而硬阈值函数可以保留信号的具体特征，但是在信号的平滑方面略有欠缺。一般而言，在小波去噪的过程中，软阈值函数比较常见。

3 自由分布式 FDR 假设检验阈值算法

3.1 自由分布式 FDR 假设检验方法

在统计学中，当有 m 个待检验零假设 $H_1, H_2, H_3, \cdots, H_m$ 被同时检验时，假设被拒绝的零假设的数目为 R，被错误拒绝的零假设数目为 V，那么被错误拒绝的零假设占所有被拒绝的零假设的比例为 $Q=V/R$。

定义随机变量 Q 的数学期望为：

$$E(Q)=E(V/R) \tag{11}$$

在特殊情况下，当没有假设被拒绝，即 R=0 时，Q=0。在多假设检验中要控制 $E(Q)$ 在一个较低的错误率水平，即 $E(Q)\le q$。

自从 Abramovich 和 Benjamini 于 1995 年提出在多假设检验中引入 FDR 准则后，多种基于 FDR 准则的假设检验方法相继出现，但是它们均建立在对检验假设的联合概率作一定假定的基础上[8-18]。唯一一种不需要对检验假设的联合概率作出假定的 FDR 假设检验方法就是自由分布式 FDR 假设检验方法，其检验流程如下：

设 $p_1, p_2, \cdots, p_m$ 是一组待检验零假设 $H_1, H_2, H_3, \cdots, H_m$ 所对应的门限概率，把它们按从小到大的顺序进行排列，可得：$p_{(1)}\le p_{(2)}\le\cdots\le p_{(m)}$。

$$k=\max\left\{i: p_{(i)}\le\frac{i}{m(1+1/2+\cdots 1/m)}q\right\} \tag{12}$$

那么待检验零假设 $H_1, H_2, H_3, \cdots, H_k$ 将被拒绝，这样上述流程就把 $E(Q)$ 控制在 q 以下。此检验过程从最大的 $p_{(m)}$ 开始，逐步向最小的 $p_{(1)}$ 进行，直至找

到第一个满足式(12)的 $p_{(i)}$，此时，对应的下标 i 就是要找的 k。

3.2 自由分布式 FDR 假设检验阈值算法

从统计学的观点解释，自由分布式 FDR 假设检验阈值算法就是对小波变换后得到的每一个小波系数进行检验，然后仅保留显著的系数。考虑对一组待检验零假设 $H_k(k=1,2,\cdots,m):\tilde{\theta}_{jk}=0$ 同时进行检验(这里将系数分为 0 和非 0 是对实际情况的理想化，而实际上信号的小波系数很少有正好为 0 的，它们仅仅是很小而已)。若在检验中共有 R 个小波系数检验失败，在这些检验失败的系数中，设由原始信号得到的小波系数的数目为 S，由噪声得到的小波系数的数目为 V，则小波系数被错误保留的比例为 $Q=V/R$。为了最大限度地减少由噪声引起的小波系数，可以通过控制 Q 的数学期望 $E(Q)$ 来实现。FDR 算法的具体步骤如下：

(1) 计算每个带噪信号小波系数 $\tilde{\theta}_{jk}$ 对应的门限概率 p_{jk}，检验 $H_k:\tilde{\theta}_{jk}=0$。

$$p_{jk}=2(1-\Phi(\frac{|\tilde{\theta}_{jk}|}{\sigma})) \tag{13}$$

式中，Φ 为高斯分布函数，σ 为高斯白噪声的标准方差。

(2) 对 p_{jk} 按从小到大进行排序，即：$p_{(1)}\le p_{(2)}\le p_{(3)}\le\cdots\le p_{(m)}$。

(3) 从 $i=1$ 开始，设满足不等式(14)的最大的 i 为 k：

$$p_{(i)}\le\frac{i}{m(1+1/2+\cdots 1/m)}q \tag{14}$$

计算出：

$$\lambda_k=\sigma\Phi^{-1}(1-\frac{p_{(k)}}{2}) \tag{15}$$

(4) 把 λ_k 作为 FDR 算法确定的阈值，选择软阈值函数或硬阈值函数对 $\tilde{\theta}_{jk}$ 进行阈值化处理。

4 仿真实验与结果分析

在信号长度 $N=1024$、高斯白噪声的标准方差 $\sigma=1$ 时，由式(8)可以求出通用阈值 $\lambda=3.7233$。当错误率 q 分别取 0.05、0.1、0.15 时，则分别有 5、10、15 个错误系数被保留，由式(15)可以求出 FDR 算法所确定的阈值分别为 $\lambda_5=3.9935$、$\lambda_{10}=3.8261$、$\lambda_{15}=3.7250$。调整 q 的取值，使被保留的错误系数的个数 $K=1,2,\cdots,20$ 时，通用阈值算法和 FDR 算法所确定的阈值如图 1 所示。从图 1 可以看出，随着 k 的增加，通用阈值保持不变而 FDR 算法所确定的阈值逐渐减小。由此可见，FDR 算法不依赖信号长度，可以通过控制不同的错误率 q 来确定不同的阈值，它克服了通用阈值依赖信号长度的这一缺陷。

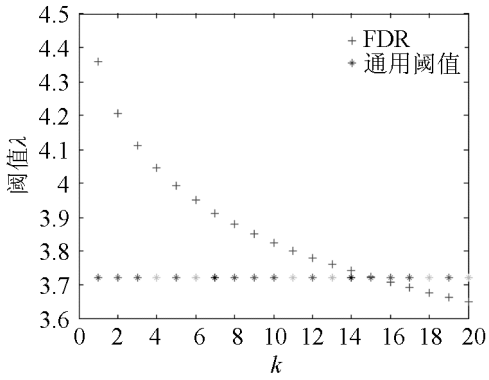

图 1 取 k 个系数时两种算法所确定的阈值

Fig.1 Thresholds for k coefficients by two algorithms

在 Matlab 软件中选取 sym4 小波函数来进行通用阈值算法和 FDR 算法的信号增强仿真实验。其中 FDR 算法的错判率分别取 $q=0.05$ (FDR05)和 $q=0.01$(FDR01)，采用 Donoho 为验证小波阈值信号增强算法而构造的 Blocks、Bumps、HeavySine、Doppler 四种理想信号作为本实验的原始信号，它们分别由 wnoise 函数在 Matlab 中自动生成。高斯白噪声的标准方差 $\sigma=1$，信号长度 $N=1024$。在原始信号中加入高斯白噪声，生成初始信噪比(SNR)分别为 1dB、5dB、10dB、15dB 的带噪信号。接着在 Matlab 软件中用 sym4 小波函数分别对上述的四种带噪信号进行小波变换，依次用通用阈值算法和 FDR05、FDR01 算法来确定阈值，采用软阈值函数对小波系数做统一的阈值化处理，最后通过小波逆变换，重构原始信号。以最小相对均方误差(Minimal Relative Mean Square Error, MRMSE)作为评价指标。最小相对均方误差的定义为：$\mathrm{MRMSE}=\min(\mathrm{MSE}_i)/\mathrm{MSE}_i$ $(i=1,2,3)$，其中，均方误差(Mean Square Error, MSE)的定义为：$\mathrm{MSE}=\frac{1}{N}\sum_{i=1}^{N}|f_i-\hat{f}_i|^2$，其中，$f_i$ 为原始信号，$\hat{f}_i$ 为增强后的信号。MRMSE 的统计结果如表 1 所示。从表 1 可以看出，FDR 算法的信号增强效果相对于通用阈值算法而言显得更加稳定。在 SNR=10dB 时，原始的 HeavySine 信号如图 2 所示，用通用阈值算法和 FDR05、FDR01 算法对图 3 所示的带噪的 HeavySine 信号进行处理，增强后的波形如图 4~6 所示。通过比较可以看出，FDR 算法明显优于通用阈值算法，而且随着错判率 q 的减小，FDR01 在信号突变处的增强效果略优于 FDR05 的增强效果。

表 1　两种阈值算法的最小相对均方误差统计结果
Table 1　MRMSE statistical results by two threshold algorithms

SNR/dB	MRMSE			
	测试信号	通用阈值法	FDR05	FDR01
1	Blocks	0.7320	0.7858	1.0000
	Bumps	1.0000	0.7903	0.8365
	HeavySine	0.5011	1.0000	0.8011
	Doppler	1.0000	0.9512	0.7899
5	Blocks	1.0000	0.7768	0.8501
	Bumps	0.7011	1.0000	0.8156
	HeavySine	0.6991	0.8502	1.0000
	Doppler	0.8547	0.7968	1.0000
10	Blocks	1.0000	0.7759	0.8482
	Bumps	0.8874	1.0000	0.8763
	HeavySine	0.6191	1.0000	0.8508
	Doppler	1.0000	0.8166	0.8522
15	Blocks	0.8330	0.7953	1.0000
	Bumps	0.8136	1.0000	0.9419
	HeavySine	1.0000	0.8602	0.9521
	Doppler	0.8500	1.0000	0.8576

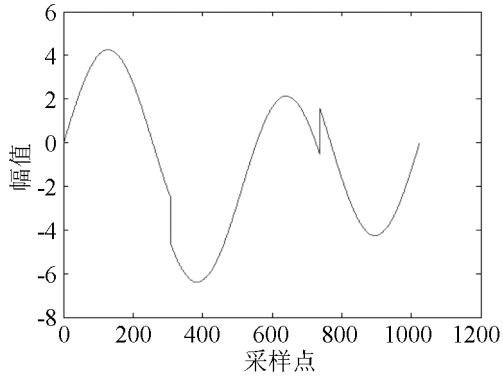

图 2　原始的 HeavySine 信号
Fig.2　The original HeavySine signal

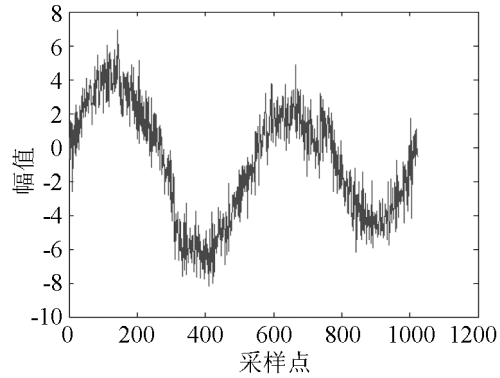

图 3　带噪的 HeavySine 信号
Fig.3　The noisy HeavySine signal

5　结 论

阈值的设定直接关系到小波阈值增强算法的增强效果。本文应用统计学中的自由分布式 FDR 假设检验算法来设定小波阈值，得到一种全新的阈值选取算法。与 Donoho 提出的通用阈值相比，该算法具有不依赖于信号长度，阈值选取灵活，并且对信号及噪声的变化具有较强的适应性等优点，仿真实验显示该算法的语音增强效果明显优于通用阈值。因此，拓展小波理论，研究小波阈值增强算法，特别是阈值的设定和阈值函数的选取，无论在学术研究上还是在实际的应用中都具有十分重要意义。

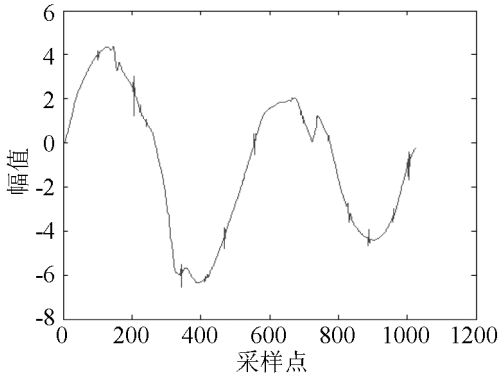

图 4　通用阈值增强后的 HeavySine 信号
Fig.4　The enhancement HeavySine signal by universal threshold

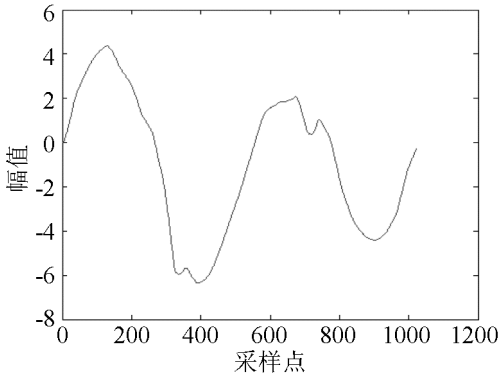

图 5　FDR05 增强后的 HeavySine 信号
Fig.5　The enhancement HeavySine signal by FDR05

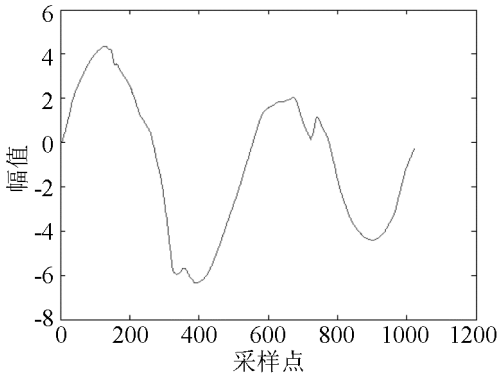

图 6　FDR01 增强后的 HeavySine 信号
Fig.6　The enhancement HeavySine signal by FDR01

参　考　文　献

[1] WANG X, ZHAO C H, RONG J G. Wavelet thresholding denoising based on simplex-simulated annealing algorithm[A]. ICIEA 2007. 2nd IEEE Conference on Industrial Electronics and Applications[C]. 2007: 2397-2402.

[2] ZHAO T, YU Y. Threshold denoising analysis of machinery vibrating signal based on wavelet transform[A]. ICEMI 2007. 8th

International Conference on Electronic Measurement and Instruments[C]. 2007: 31041-31044.

[3] Ferrando S E, Pyke R. Ideal denoising for signals in sub-Gaussian noise[J]. Applied and Computational Harmonic Analysis, 2008, **24**(1): 1-13.

[4] Poornachandra S. Wavelet-based denoising using subband dependent threshold for ECG signals[J]. Digital Signal Processing, 2008, **18**(1): 49-55.

[5] Johnson M T, YUAN X L, REN Y. Speech signal enhancement through adaptive wavelet thresholding[J]. Speech Communication, 2007, **49**(2): 123-133.

[6] Azzalini A, Farge M, Schneider K. Nonlinear wavelet thresholding: A recursive method to determine the optimal denoising threshold[J]. Applied and Computational Harmonic Analysis, 2005, **18**(2): 177-185.

[7] Donoho D L. De-noising by soft-thresholding[J]. IEEE Transactions on Information Theory, 1995, **41**(3): 613-627.

[8] Abramovich F, Benjamini Y. Thresholding of wavelet coefficients as multiple hypotheses testing procedure[Z]. Wavelets and Statistics, Lecture Notes in Statistics 103, New York: Springer-Verlag, 1995: 5-14.

[9] Abramovich F, Benjamini Y, Donoho D L, et al. Adapting to unknown sparsity by controlling the false discovery rate[J]. Annals of Statistics, 2006, **34**(2): 584-653.

[10] GUO W, RAO M B. On control of the false discovery rate under no assumption of dependency[J]. Journal of Statistical Planning and Inference, 2008, **138**(10): 3176-3188.

[11] Ge Y C, Sealfon S C, Tseng C H, et al. A Holm-type procedure controlling the false discovery rate[J]. Statistics & Probability Letters, 2007, **77(**18): 1756-1762.

[12] Tsui K W, Tang S J. Simultaneous testing of multiple hypotheses using generalized p-values[J]. Statistics & Probability Letters[J]. 2007, **77**(12): 1362-1370.

[13] Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics [J]. Journal of Educational and Behavioral Statistics, 2000, **25**(1): 60-83.

[14] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency[J]. Annals of Statistics, 2001, **29**(4): 1165-1188.

[15] Tadesse M G, Ibrahim J G, Vannucci M, et al. Wavelet thresholding with Bayesian false discovery rate control[J]. Biometrics, 2005, **61**(1): 25-35.

[16] Storey J D, Taylor J E, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach[J]. Journal of the Royal Statistical Society. Series B, Statistical Methodology, 2004, **66**(1): 187-205.

[17] Storey J D. A direct approach to false discovery rates[J]. Journal of the Royal Statistical Society. Series B, Methodological, 2002, **64**(3): 479-498.

[18] Sarkar S K. Some results on false discovery rate in stepwise multiple testing procedures[J]. Annals of Statistics, 2002, **30**(1): 239-257.