中国科学 : 信息科学 2010 年 第 40 卷 第 1 期 : 33 ~ 40
www.scichina.com info.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS
论 文

基于稀疏表示的小波去噪

赵瑞珍[①②③*], 刘晓宇[②③], LI ChingChung [③], SCLABASSI Robert J[②③], 孙民贵[②③]

① 北京交通大学信息科学研究所, 北京 100044
② Department of Neurosurgery, University of Pittsburgh, Pittsburgh, PA 15261, USA
③ Department of Electrical & Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA
* 通信作者. E-mail: rzhzhao@bjtu.edu.cn

收稿日期: 2008-11-18; 接受日期: 2009-01-14
美国卫生部项目 (批准号: U01HL91736)、教育部留学回国人员科研启动基金 (批准号: [2009]1341) 和国家重点基础研究发展计划 (批准号: 2007AA01Z175) 资助项目

摘要 小波阈值去噪是信号与图像去噪中的有效方法, 然而, 该方法采用逐点处理的方式, 未用到小波系数的整体结构特性. 文中提出一种新的小波去噪方法, 采用了新近发展起来的稀疏表示工具, 通过在一定条件下最小化非零小波系数的个数对原小波系数进行估计, 从而将去噪转化为一个最优化问题. 证明了该优化问题的解可以惟一获得, 并且该解是干净小波系数的一个无偏估计值. 文中提出了一种求解该问题的方法, 该方法至少能求得一个局部最优解. 实验结果表明此方法对多数实际信号尤其是低信噪比信号是有效的.

关键词
信号处理
去噪
稀疏表示
小波变换

1 引言

目前, 小波去噪的方法大致可以分为 3 类: 第 1 类方法是 Mallat 等 [1] 提出的基于小波变换模极大值原理的去噪方法, 其原理是信号和噪声在小波变换域内随尺度的变化具有不同的传播特性; 第 2 类方法是通过计算相邻尺度间各小波系数的相关性 [2], 并根据其大小决定小波系数的取舍; 第 3 类方法是 Donoho 等 [3] 提出的阈值方法, 在众多小波系数中, 把绝对值较小的系数置为零, 而让绝对值较大的系数保留或收缩 (shrinkage), 得到小波系数的估计值, 从而进行信号重构, 阈值方法又分为硬阈值和软阈值方法两种.

在 3 类小波去噪方法中, 阈值方法由于其简单有效, 应用最为广泛. 然而该方法可能在不连续区域出现 Gibbs 现象 [4], 且当信噪比较低时其去噪性能明显下降; 另外, 阈值去噪方法处理小波系数时采用逐系数处理的方式, 忽略了小波系数的整体结构特性. 在过去几年内出现了许多阈值去噪的改进形式. 例如, Cai 等 [5] 利用小波系数的邻域特性提出 NeighShrink 方法处理一维信号; Chen 等 [6] 将其推广至二维情况用于图像去噪. 近年来在数学和工程领域同时兴起的压缩传感与稀疏表示理论给信号去噪提供了有力的工具 [7,8]. 目前已有一些关于含噪压缩测量的重构算法 [9~12], 然而这些算法针对的噪声是压缩传感过程中引入的噪声. 本文提出一种基于稀疏表示的小波去噪方法, 首先对含噪信号

引用格式: 赵瑞珍, 刘晓宇, Li C C, 等. 基于稀疏表示的小波去噪. 中国科学: 信息科学, 2010, 40: 33–40

进行小波变换, 并得到一组小波系数, 然后对这些系数建立稀疏表示模型. 不同于阈值方法中逐系数处理的方式, 该方法引入随机测量矩阵对小波系数进行变换, 并通过在一定条件下最小化非零小波系数个数的方法获得理想的干净小波系数. 最后通过小波反变换获得去噪之后的信号. 本文方法可获得相当好的去噪效果.

本文安排如下: 第 2 节提出基于稀疏表示的去噪模型, 并讨论了该模型解的存在性和惟一性. 第 3 节讨论了在一定的参数条件下, 至少可获得该模型的局部最优解. 第 4 节给出新的去噪算法的大致步骤. 第 5，6 节分别是实验结果和结论.

2 基于稀疏表示的去噪模型

设有如下含噪信号

$$f(k) = s(k) + n(k), \tag{1}$$

其中 $f(k)$ 和 $s(k)$ 分别是观测信号和干净无噪信号, $n(k)$ 是随机噪声且服从正态分布 $N(0, \sigma^2)$. 设 $w_{j,k}$ 表示信号 $f(k)$ 对应的小波系数, $j = 1 \sim J$ 为尺度因子, $k = 1 \sim N$ 表示时间 (位置). Donoho 提出的小波去噪算法分为软阈值和硬阈值方法两种, $\hat{w}_{j,k}$ 表示通过阈值方法估计得到的小波系数.

软阈值:

$$\hat{w}_{j,k} = \begin{cases} \operatorname{sign}(w_{j,k}) \cdot (|w_{j,k}| - \lambda), & |w_{j,k}| \geqslant \lambda, \\ 0, & |w_{j,k}| < \lambda; \end{cases} \tag{2}$$

硬阈值:

$$\hat{w}_{j,k} = \begin{cases} w_{j,k}, & |w_{j,k}| \geqslant \lambda, \\ 0, & |w_{j,k}| < \lambda. \end{cases} \tag{3}$$

由 (2) 和 (3) 式明显看到, 该操作属于逐点处理方式, 未用到小波系数的整体结构特性.

对于多数信号而言, 只要小波基选择合适, 对其进行小波变换后得到的小波系数是稀疏的, 即仅有少数大幅值系数. 然而, 若信号被噪声污染, 则小波系数的稀疏性将大大降低. 因此, 可以考虑采用恢复小波系数的稀疏性达到去噪的目的. 我们引入稀疏表示技术 [13], 旨在真实估计重要小波系数但不丢失重要信息, 将去噪归结为如下最优化问题.

不失一般性, 略去小波系数 $w_{j,k}$ 的下标 j 和 k, 令 $\boldsymbol{w}$ 表示给定尺度上大小为 $N \times 1$ 的列向量. $\boldsymbol{\Phi}_{M \times N}$($M$ 小于 N) 是稀疏表示中的随机测量矩阵. 作为一个基本要求, $\boldsymbol{\Phi}$ 必须满足一致不确定原理 (uniform uncertainty principle, UUP)[14], 即对于任意 S– 稀疏向量 $\boldsymbol{x}$, 如果

$$0.8\frac{M}{N}\|\boldsymbol{x}\|_2^2 \leqslant \|\boldsymbol{\Phi x}\|_2^2 \leqslant 1.2\frac{M}{N}\|\boldsymbol{x}\|_2^2, \tag{4}$$

则称 $\boldsymbol{\Phi}_{M \times N}$ 满足集合大小为 S 的 UUP 原理, 其中 $S \leqslant M \log N$.

设 $\boldsymbol{y} = \boldsymbol{\Phi w}$, 我们可以通过求解下列问题获得最稀疏的系数向量 $\hat{\boldsymbol{w}}$,

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_0 \quad \text{s.t. } \|\boldsymbol{y} - \boldsymbol{\Phi z}\|_2 \leqslant \varepsilon, \tag{5}$$

其中 $\|\cdot\|_0$ 表示零范数, 即非零元素的个数 [7]. 一般来讲, (5) 式很难求解, 因为在组合数学中它属于 NP 难问题. 然而, 在去噪应用中, 我们可以找到一种求解该问题的简便方式. 将 (5) 式修改为如下的

形式 [12]:

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{z}}(\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{z}\|_2^2 + \gamma\|\boldsymbol{z}\|_0). \tag{6}$$

下面我们将验证在某种条件下, (6) 式的解在统计意义下是惟一的, 且是无偏的, 其均值恰好和理想的小波系数相同.

令 $\bar{\boldsymbol{w}}$ 表示无噪小波系数, 通过加性假设, 观测信号的小波系数可表示为 $\boldsymbol{w} = \bar{\boldsymbol{w}} + \boldsymbol{w}_n$, 其中 $\boldsymbol{w}_n$ 是随机噪声 $n(k)$ 对应的小波系数. 由于 $n(k) \sim N(0, \sigma^2)$, $\boldsymbol{w}_n$ 也服从零均值 Gauss 分布. 因此我们有

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{w} = \boldsymbol{\Phi}(\bar{\boldsymbol{w}} + \boldsymbol{w}_n) = \boldsymbol{\Phi}\bar{\boldsymbol{w}} + \boldsymbol{\Phi}\boldsymbol{w}_n. \tag{7}$$

记

$$g(\boldsymbol{z}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{z}\|_2^2 + \gamma\|\boldsymbol{z}\|_0, \tag{8}$$

我们将证明对于任意 $\boldsymbol{u} \neq \bar{\boldsymbol{w}}$, 有 $\mathrm{E}(g(\bar{\boldsymbol{w}})) < \mathrm{E}(g(\boldsymbol{u}))$, 即 $\mathrm{E}(\hat{\boldsymbol{w}}) = \bar{\boldsymbol{w}}$.

设 $\boldsymbol{\Phi}$ 表示随机矩阵, 其元素服从 Gauss 分布 $N(0, \sigma_{\boldsymbol{\Phi}}^2)$. 联合 (7) 和 (8) 式, 我们考虑

$$\begin{aligned} g(\boldsymbol{u}) - g(\bar{\boldsymbol{w}}) &= (\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2 - \|\boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{w}}\|_2^2) + \gamma(\|\boldsymbol{u}\|_0\|\bar{\boldsymbol{w}}\|_0) \\ &= (\|\boldsymbol{\Phi}(\bar{\boldsymbol{w}} + \boldsymbol{w}_n - \boldsymbol{u})\|_2^2 - \|\boldsymbol{\Phi}\boldsymbol{w}_n\|_2^2) + \gamma(\|\boldsymbol{u}\|_0 - \|\bar{\boldsymbol{w}}\|_0) \\ &= (\|\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - \boldsymbol{u}) + \boldsymbol{\Phi}\boldsymbol{w}_n\|_2^2 - \|\boldsymbol{\Phi}\boldsymbol{w}_n\|_2^2) + \gamma(\|\boldsymbol{u}\|_0 - \|\bar{\boldsymbol{w}}\|_0) \\ &= (\|\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - \boldsymbol{u})\|_2^2 + 2[\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - u)]^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{w}_n) + \gamma(\|\boldsymbol{u}\|_0 - \|\bar{\boldsymbol{w}}\|_0). \end{aligned} \tag{9}$$

假设 $\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - \boldsymbol{u})$ 和 $\boldsymbol{\Phi}\boldsymbol{w}_n$ 独立, 则易得

$$\mathrm{E}(g(\boldsymbol{u}) - g(\bar{\boldsymbol{w}})) = \mathrm{E}(\|\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - \boldsymbol{u})\|_2^2) + \gamma\mathrm{E}(\|\boldsymbol{u}\|_0 - \|\bar{\boldsymbol{w}}\|_0). \tag{10}$$

下面我们将分 3 种情况对 (10) 式进行讨论:

i) 当 $\|\boldsymbol{u}\|_0 > \|\bar{\boldsymbol{w}}\|_0$ 时, $\mathrm{E}(g(\boldsymbol{u}) - g(\bar{\boldsymbol{w}})) > 0$ 显然成立.

ii) 当 $\|\boldsymbol{u}\|_0 = \|\bar{\boldsymbol{w}}\|_0$ 时, 假设 $\|\boldsymbol{u}\|_0 = \|\bar{\boldsymbol{w}}\|_0 = K$. 记 $\boldsymbol{v} = \bar{\boldsymbol{w}} - \boldsymbol{u}$, 于是 $\|\boldsymbol{v}\|_0 \leqslant 2K$. 令 $\Gamma = \{l | v_l \neq 0\}$, 显然集合 Γ 的大小满足 $|\Gamma| \leqslant 2K$. 因此, 我们可以将 $\Phi(\bar{\boldsymbol{w}} - \boldsymbol{u})$ 重写为 $\boldsymbol{\Phi}\boldsymbol{v} = \sum_{l\in\Gamma} v_l\boldsymbol{\Phi}_l$, 其中 $\boldsymbol{\Phi}_l$ 表示矩阵 $\boldsymbol{\Phi}_{M\times N}$ 的列向量. 由于 $\boldsymbol{\Phi}$ 满足 UUP 特性, 所以 $\boldsymbol{\Phi}$ 的任意 M 列均是近似线性独立的. 因此, 若 $2K \leqslant M$, 由于对于 $l \in \Gamma$ 有 $v_l \neq 0$, 所以可得 $\sum_{l\in\Gamma} v_l\boldsymbol{\Phi}_l \neq \boldsymbol{0}$ 几乎总是成立. 因此, (10) 式右侧的第 1 项大于 0, 从而有 $\mathrm{E}(g(\boldsymbol{u}) - g(\bar{\boldsymbol{w}})) > 0$.

iii) 当 $\|\boldsymbol{u}\|_0 < \|\bar{\boldsymbol{w}}\|_0$ 时, 假设 $\|\boldsymbol{u}\|_0 = \|\bar{\boldsymbol{w}}\|_0 - \eta$, 则有

$$\mathrm{E}(g(\boldsymbol{u}) - g(\bar{\boldsymbol{w}})) = \mathrm{E}(\|\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - \boldsymbol{u})\|_2^2) - \gamma\eta. \tag{11}$$

同样记 $\boldsymbol{v} = \bar{\boldsymbol{w}} - \boldsymbol{u}$ 和 $\Gamma = \{l | v_l \neq 0\}$, 则有 $\eta \leqslant \|\boldsymbol{v}\|_0 = |\Gamma| \leqslant 2K - \eta$. 因此

$$\|\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - \boldsymbol{u})\|_2^2 = \|\boldsymbol{\Phi}\boldsymbol{v}\|_2^2 = \left\|\sum_{l\in\Gamma} v_l\boldsymbol{\Phi}_l\right\|_2^2 = \sum_{i=1}^{M}\left(\sum_{l\in\Gamma} v_l\boldsymbol{\Phi}_{il}\right)^2. \tag{12}$$

根据随机矩阵 $\boldsymbol{\Phi}$ 的统计特性, 我们有 $\mathrm{E}(\boldsymbol{\Phi}_{i,l}) = 0$, $\mathrm{E}(\boldsymbol{\Phi}_{i,l}^2) = D(\boldsymbol{\Phi}_{i,l}) = \sigma_{\boldsymbol{\Phi}}^2$, 且对于 $l \neq \tilde{l}$ 有 $E(\boldsymbol{\Phi}_{i,l}\boldsymbol{\Phi}_{i,\tilde{l}}) = 0$. 因此, (12) 式的期望为

$$\mathrm{E}(\|\boldsymbol{\Phi}(\bar{\boldsymbol{w}} - \boldsymbol{u})\|_2^2) = M\sigma_{\boldsymbol{\Phi}}^2\sum_{l\in\Gamma} v_l^2 = M\sigma_{\boldsymbol{\Phi}}^2\|\boldsymbol{v}\|_2^2. \tag{13}$$

结合 (11) 式, 如果 $\boldsymbol{v}$ 满足下列条件

$$\|\boldsymbol{v}\|_2^2 > \frac{\gamma\eta}{M\sigma_{\boldsymbol{\Phi}}^2}, \tag{14}$$

则有 $\mathrm{E}(g(\boldsymbol{u}) - g(\bar{\boldsymbol{w}})) > 0$. 由于 $\|\boldsymbol{u}\|_0 = \|\bar{\boldsymbol{w}}\|_0 - \eta$, 在向量 $\boldsymbol{v}$ 中至少存在 η 个非零系数, 使得 $v_l = \bar{w}_l$ $(l \in \Gamma)$. 另外, 因为 M 通常远大于 1, 并且 $\sigma_{\boldsymbol{\Phi}}$ 的值可在构建矩阵 $\boldsymbol{\Phi}$ 时进行适当选择, 所以 $\boldsymbol{v}$ 中的每一个非零元素在实际中均可大于 $\gamma/(\sqrt{M}\sigma_{\boldsymbol{\Phi}})$. 由此可得, (14) 式容易满足且在实际中是一个弱条件.

综上所述, 对于任意的 $\boldsymbol{u} \neq \bar{\boldsymbol{w}}$, 我们有 $E(g(\boldsymbol{u}) - g(\bar{\boldsymbol{w}})) > 0$. 由于 $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{z}} g(\boldsymbol{z})$, 所以有 $E(\hat{\boldsymbol{w}}) = \bar{\boldsymbol{w}}$.

3 最优化问题的解

令 $g_1(\boldsymbol{z})$ 表示 (8) 式右侧的第 1 项, 通过对 $g_1(\boldsymbol{z})$ 求偏导并令其等于零可得到最小化 $g_1(\boldsymbol{z})$ 的必要条件:

$$\frac{\partial g_1(\boldsymbol{z})}{\partial \boldsymbol{z}} = -2\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{z}) = \boldsymbol{0}. \tag{15}$$

于是, 求上述问题可采用最速下降法

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} + \mu\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{w}^{(t)}), \tag{16}$$

其中, $\mu > 0$ 是一个待定常数. (16) 式可重写为

$$\boldsymbol{w}^{(t+1)} = \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} + (\boldsymbol{I} - \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})\boldsymbol{w}^{(t)}. \tag{17}$$

因此, (17) 式收敛的条件是 $(\boldsymbol{I} - \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})$ 应为一线性压缩算子, 即

$$0 < \|\boldsymbol{I} - \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\|_2 < 1. \tag{18}$$

上述算法收敛的另一条件是线性算子 $(\boldsymbol{I} - \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})$ 的特征值满足下列充分条件 [15], 该条件强于 (18) 式

$$0 < \mathrm{eig}(\boldsymbol{I} - \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}) < 1. \tag{19}$$

为了说明该条件的有效性, 我们采用 b 表示矩阵 $\boldsymbol{I} - \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$ 的任意特征值. 那么, 对于任意特征向量 $\boldsymbol{\alpha} \neq \boldsymbol{0}$, 有 $(\boldsymbol{I} - \mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})\boldsymbol{\alpha} = b\boldsymbol{\alpha}$ 或 $\mu\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{\alpha} = (1-b)\boldsymbol{\alpha}$. 由于矩阵 $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$ 依概率 1 正定, 因此 b 应当小于 1, 从而 (19) 式的上界容易满足. 为了满足其下界, 需要对于任意的 μ, 有 $\|\sqrt{\mu}\boldsymbol{\Phi}\|_2^2 < 1$. 不幸的是, 满足 UUP 特性的矩阵 $\boldsymbol{\Phi}$ 并不能自动满足条件 $\|\sqrt{\mu}\boldsymbol{\Phi}\|_2^2 < 1$. 然而, 若定义 s 为矩阵 $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$ 的最大特征值, 则对于任意向量 $\boldsymbol{\alpha} \neq \boldsymbol{0}$, 我们有

$$\|\sqrt{\mu}\boldsymbol{\Phi}\|_2^2 = \mu \max_{\boldsymbol{\alpha}} \frac{\|\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2}{\|\boldsymbol{\alpha}\|_2^2} = \mu \max_{\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_2^2} = \mu\frac{\boldsymbol{\alpha}^{\mathrm{T}}(s\boldsymbol{\alpha})}{\|\boldsymbol{\alpha}\|_2^2} = \mu s. \tag{20}$$

因此, 为了保证迭代算法的收敛性, (16) 式中的 μ 应选为 $\mu < \frac{1}{s}$.

然而, 由于迭代公式 (16) 并未考虑 $\gamma\|\boldsymbol{w}\|_0$ 这一项, 因此上述算法不一定能收敛到 (6) 式的最小解. 为了最小化 $\gamma\|\boldsymbol{w}\|_0$, 我们在 (16) 式中增加如下阈值操作:

$$\boldsymbol{w}^{(t+1)} = H_\delta(\boldsymbol{w}^{(t)} + \mu\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{w}^{(t)})), \tag{21}$$

其中 H_δ 是硬阈值算子

$$H_\delta(w_i)=\begin{cases} w_i, & \text{当 } |w_i|\geqslant\delta,\\ 0, & \text{其他,}\end{cases}\qquad i=1\sim N, \tag{22}$$

其中 δ 是一可选阈值. 该算法中的硬阈值步骤保证了最小化 $\boldsymbol{w}$ 中的非零系数个数得以实现.

下面我们将验证迭代算法 (21) 的不动点 $\hat{\boldsymbol{w}}$ 就是 (6) 式的最小解, 即存在 $\xi>0$, 使得对于任意满足 $|\Delta w_i|<\xi$ 的 $\Delta\boldsymbol{w}$, $g(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w})>g(\hat{\boldsymbol{w}})$ 总成立.

我们引入如下新的辅助目标函数 [16]

$$G(\boldsymbol{w},\boldsymbol{u})=g(\boldsymbol{w})-\mu\|\boldsymbol{\Phi w}-\boldsymbol{\Phi u}\|_2^2+\|\boldsymbol{w}-\boldsymbol{u}\|_2^2=\|\boldsymbol{y}-\boldsymbol{\Phi w}\|_2^2+\gamma\|\boldsymbol{w}\|_0-\mu\|\boldsymbol{\Phi w}-\boldsymbol{\Phi u}\|_2^2+\|\boldsymbol{w}-\boldsymbol{u}\|_2^2, \tag{23}$$

若 $\|\sqrt{\mu}\boldsymbol{\Phi}\|_2^2<1$, 显然有 $G(\boldsymbol{w},\boldsymbol{u})\geqslant G(\boldsymbol{w},\boldsymbol{w})=g(\boldsymbol{w})$. 从 (23) 式还可得到

$$g(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w})=G(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w},\widehat{\boldsymbol{w}})-\|\Delta\boldsymbol{w}\|_2^2+\mu\|\boldsymbol{\Phi}\Delta\boldsymbol{w}\|_2^2. \tag{24}$$

(23) 式可重写为

$$G(\boldsymbol{w},\boldsymbol{u})=\sum_i[w_i^2-2w_i(u_i+\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{y}-\mu\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi u})+\gamma|w_i|^0]+(1-\mu)\|\boldsymbol{\Phi w}\|_2^2+\|\boldsymbol{y}\|_2^2+\|\boldsymbol{u}\|_2^2-\mu\|\boldsymbol{\Phi u}\|_2^2, \tag{25}$$

其中, 若 $w_i\neq0$, 则 $|w_i|^0$ 为 1; 否则 $|w_i|^0$ 为 0. 利用 (25) 式, 我们考虑

$$\begin{aligned}G(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w},\hat{\boldsymbol{w}})-G(\hat{\boldsymbol{w}},\hat{\boldsymbol{w}})=&\sum_i[(\hat{w}_i+\Delta w_i)^2-2(\hat{w}_i+\Delta w_i)(\hat{w}_i+\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{y}-\mu\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}\hat{\boldsymbol{w}})+\gamma|\hat{w}_i+\Delta w_i|^0]\\&-\sum_i[\hat{w}_i^2-2\hat{w}_i(\hat{w}_i+\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{y}-\mu\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}\hat{\boldsymbol{w}})+\gamma|\hat{w}_i|^0]\\&+(1-\mu)(\|\boldsymbol{\Phi}(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w})\|_2^2-\|\boldsymbol{\Phi}\hat{\boldsymbol{w}}\|_2^2)\\=&\sum_i[(\Delta w_i)^2-2\Delta w_i\boldsymbol{\Phi}_i^{\mathrm{T}}(\boldsymbol{y}-\mu\boldsymbol{\Phi}\hat{\boldsymbol{w}})+\gamma(|\hat{w}_i+\Delta w_i|^0-|\hat{w}_i|^0)]\\&+(1-\mu)\sum_i2\Delta w_i\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}\hat{\boldsymbol{w}}+(1-\mu)\|\boldsymbol{\Phi}\Delta\boldsymbol{w}\|_2^2\\=&\|\Delta\boldsymbol{w}\|_2^2+(1-\mu)\|\boldsymbol{\Phi}\Delta\boldsymbol{w}\|_2^2+\sum_i[-2\Delta w_i\boldsymbol{\Phi}_i^{\mathrm{T}}(y-\boldsymbol{\Phi}\hat{\boldsymbol{w}})\\&+\gamma(|\hat{w}_i+\Delta w_i|^0-|\hat{w}_i|^0)].\end{aligned}$$

为讨论方便, 将上式中的求和项进行拆分, 在集合 $\Gamma_0=\{i|\hat{w}_i=0\}$ 和 $\Gamma_1=\{i|\hat{w}_i\neq0\}$ 上分别进行求和. 对于 $i\in\Gamma_1$, 若 $\Delta w_i\neq-\hat{w}_i$, 我们得到

$$\begin{aligned}G(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w},\hat{\boldsymbol{w}})-G(\hat{\boldsymbol{w}},\hat{\boldsymbol{w}})=&\|\Delta\boldsymbol{w}\|_2^2+(1-\mu)\|\boldsymbol{\Phi}\Delta\boldsymbol{w}\|_2^2\\&-\sum_{\Gamma_0}[2\Delta w_i\boldsymbol{\Phi}_i^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\Phi}\hat{\boldsymbol{w}})-\gamma|\Delta w_i|^0]-\sum_{\Gamma_1}[2\Delta w_i\boldsymbol{\Phi}_i^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\Phi}\hat{\boldsymbol{w}})].\end{aligned} \tag{26}$$

根据迭代 (21) 式, 对于 $i\in\Gamma_1$, $\hat{w}_i=\hat{w}_i+\mu\boldsymbol{\Phi}_i^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\Phi}\hat{\boldsymbol{w}})$ 恒成立. 因此, 对于 $i\in\Gamma_1$ 我们有 $\boldsymbol{\Phi}_i^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\Phi}\hat{\boldsymbol{w}})=\boldsymbol{0}$, 这意味着 (26) 式中第 2 个求和项等于 0. 令 $\xi=\min_{i\in\Gamma_0}\frac{\gamma}{2|\boldsymbol{\Phi}_i^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\Phi}\hat{\boldsymbol{w}})|}$, 假设对于 $i\in\Gamma_0$, 有 $|\Delta w_i|\leqslant\xi$; 对于 $i\in\Gamma_1$, 有 $\Delta w_i\neq-\hat{w}_i$, 则易得

$$G(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w},\hat{\boldsymbol{w}})-G(\hat{\boldsymbol{w}},\hat{\boldsymbol{w}})\geqslant\|\Delta\boldsymbol{w}\|_2^2+(1-\mu)\|\boldsymbol{\Phi}\Delta\boldsymbol{w}\|_2^2. \tag{27}$$

将 (27) 式代入 (24) 式可得

$$g(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w})=G(\hat{\boldsymbol{w}}+\Delta\boldsymbol{w},\hat{\boldsymbol{w}})-\|\Delta\boldsymbol{w}\|_2^2+\mu\|\boldsymbol{\Phi}\Delta\boldsymbol{w}\|_2^2\geqslant G(\hat{\boldsymbol{w}},\hat{\boldsymbol{w}})+\|\boldsymbol{\Phi}\Delta\boldsymbol{w}\|_2^2\geqslant g(\hat{\boldsymbol{w}}). \tag{28}$$

因此, $g(\hat{\boldsymbol{w}})$ 是一个局部最小解.

4 去噪算法

由于纯净信号对应的小波系数 $\bar{w}_j$ 沿着小波分解尺度具有很强的相关性, 因此不同尺度上的非零小波系数位于相同位置邻域的可能性很大. 基于此, 如果我们已经处理得到某一尺度上的小波系数, 则完全可以利用该尺度上小波系数的位置信息去预测其他尺度上重要系数的位置分布情况. 这一特性有助于选择优化过程中每一尺度上的迭代初值, 从而可加速整个迭代算法的收敛性. 因此, 对于不同的分解尺度 $j=1\sim J$, 我们选择如下不同的迭代初值

$$\boldsymbol{\beta}_j=\begin{cases}\boldsymbol{w}_J, & j=J,\\ \hat{\boldsymbol{w}}_{j+1}, & j<J.\end{cases} \tag{29}$$

于是, 基于稀疏表示的小波去噪算法步骤可描述如下:

1) 对含噪观测信号 $f(k)$ 进行非下采样小波变换, 并得到小波系数 $\boldsymbol{w}_j$ $(j=1\sim J)$;

2) 选择合适的随机矩阵 $\boldsymbol{\Phi}_{M\times N}$, 使其满足 UUP 特性 (4) 和条件 (14). 计算 $\boldsymbol{y}_j=\boldsymbol{\Phi}\boldsymbol{w}_j$ $(j=1\sim J)$, 并令 $j=J$;

3) 令 $\boldsymbol{w}_j^{(0)}=\boldsymbol{0}$, $\boldsymbol{w}_j^{(1)}=\boldsymbol{\beta}_j$;

4) 如果 $\|\boldsymbol{w}_j^{(1)}-\boldsymbol{w}_j^{(0)}\|_2>\varepsilon$, 则令 $\boldsymbol{w}_j^{(0)}=\boldsymbol{w}_j^{(1)}$, 并利用 (21) 式计算 $\boldsymbol{w}_j^{(1)}$, 重复步骤 4); 否则, 令 $\hat{\boldsymbol{w}}_j=\boldsymbol{w}_j^{(1)}$, 转步骤 5);

5) 若 $j=1$, 转步骤 6); 否则, 令 $j=j-1$, 转 3);

6) 利用估计得到的小波系数 $\hat{\boldsymbol{w}}_j$ 进行小波反变换.

5 实验结果

我们选择 Blocks, Bumps, HeaviSine 和 Doppler 4 种标准信号作为测试信号对所提去噪算法进行试验. 测试信号的长度均为 $N=2048$, 信噪比均为 6.1320 dB. 由于其简单有效, 且具有紧支集和光滑等特性, 我们选用三次样条小波进行实验, 变换时最大分解尺度选为 $J=4$. 采用中值法估计噪声的标准差 [17]

$$\sigma=\text{Median}(|w_{1,k}|)/0.6745, \tag{30}$$

其中 $w_{1,k}$ 表示尺度 $j=1$ 上的小波系数. 实验中, 我们按照经验选取 $\varepsilon=1$, (22) 式中选取阈值 $\delta=\sigma/4$.

由表 1 可见, 相比硬阈值方法、软阈值方法以及 NeighShrink 方法, 本文提出的去噪方法 (不同的 M 值) 获得更高的信噪比增益. 另外也可看出, 随着 M 的增加, 此算法的去噪性能也随之提高. 图 1 给出了本文算法以及其他现有方法针对 Blocks 信号的去噪结果比较, 显然, 本文提出的新算法具有明显的优越性.

为说明本文算法的鲁棒性, 我们计算了该算法针对不同信号去噪后的平均信噪比变化情况. 图 2 给出了不同去噪算法针对不同信号的平均去噪性能, 可以看出, 当原信号的信噪比低于 12.5 dB 时, 本文算法取得最高的信噪比增益. 由此可得, 此算法非常适合于低信噪比信号的去噪.

表 1　不同去噪算法对信号去噪后的信噪比 SNR (dB)

	Blocks	Bumps	HeaviSine	Doppler
软阈值方法	11.4994	11.8633	21.6421	13.8678
硬阈值方法	12.7075	12.4993	21.5754	13.8762
NeighShrink 方法	15.2111	14.9829	21.6857	14.8955
本文方法 ($M = N/8$)	16.6193	17.4104	21.7276	16.2274
本文方法 ($M = N/4$)	17.0667	18.8759	21.9509	16.5522

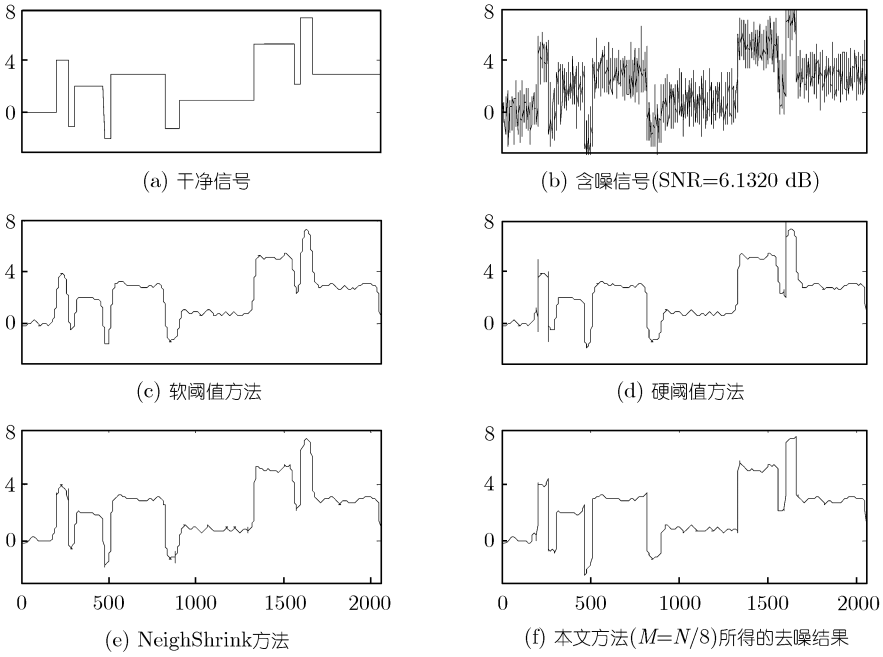

图 1　不同去噪方法对 Blocks 信号的去噪结果

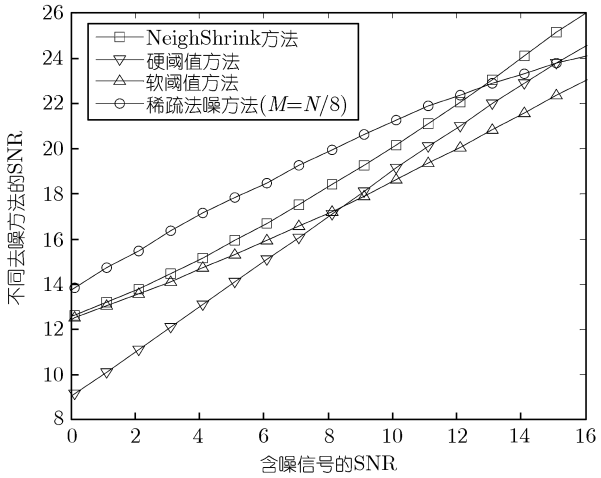

图 2　不同去噪方法所得的 SNR 比较

6 结论

尽管标准的小波阈值去噪方法简单实际, 但对低信噪比信号的去噪却无能为力. 另外, 阈值去噪采用逐点处理方式, 没有考虑小波系数的整体结构特性. 本文提出的基于小波变换和稀疏表示的新算法克服了该问题, 它通过求解最优化问题估计无噪小波系数, 对包括低信噪比信号在内的多数信号均切实有效.

参考文献

1 Mallat S, Zhong S. Characterization of signals from multiscale edges. IEEE Trans PAMI, 1992, 14: 710–732
2 Xu Y, Weaver J, Healy M, et al. Wavelet transform domain filters: a spatially selective noise filtration technique. IEEE Trans Imag Process, 1994, 3: 747–758
3 Donoho D L. De-noising by soft-thresholding. IEEE Trans Inform Theor, 1995, 41: 613–627
4 Kelly S E. Gibbs phenomenon for wavelets. Appl Comput Harmon A, 1996, 3: 72–81
5 Cai T T, Silverman B W. Incorporating information on neighboring coefficients into wavelet estimation. Indian J Stat, 2001, 63(B): 127–148
6 Chen G Y, Bui T D, Krzyzak A. Image denoising using neighboring wavelet coefficients. ICASSP, 2004, II: 917–920
7 Candès E J, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inform Theor, 2006, 52: 489–509
8 Donoho D L. Compressed sensing. IEEE Trans Inform Theor, 2006, 52: 1289–1306
9 Boufounos P, Duarte M, Baraniuk R. Sparse signal reconstruction from noisy compressive measurements using cross validation. In: Proceedings of IEEE Statistical Signal Processing Workshop (SSP) 2007. Vol 15. Madison, 2007. 358–368
10 Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Imag Process, 2006, 15: 3736–3745
11 Donoho D L, Elad M, Temlyakov V. Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans Inform Theor, 2006, 52: 6–18
12 Haupt J, Nowak R. Signal reconstruction from noisy random projections. IEEE Trans Inform Theor, 2006, 52: 4036–4048
13 Baraniuk R G. Compressive sensing. IEEE Signal Proc Mag, 2007, 24: 118–120
14 Candès E J, Tao T. Near-optimal signal recovery from random projection: universal encoding strategies? IEEE Trans Inform Theor, 2006, 52: 5406–5426
15 Blumensath T, Davies M E. Iterative thresholding for sparse approximations. J Fourier Anal Appl, 2008, 14: 629–654
16 Lange K, Hunter D R, Yang I. Optimization transfer using surrogate objective functions. J Comput Graph Stat, 2006, 9: 1–20
17 Pizurica A, Philips W, Lemahieu I, et al. A joint inter and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Trans Imag Process, 2002, 11: 545–557