

DOI:10.19322/j.cnki.issn.1006-4710.2021.02.009

PCD 算法对音频信号降噪的参数选择

何选森^{1,2}, 徐丽¹, 樊跃平¹

(1. 广州商学院 信息技术与工程学院, 广东 广州 511363; 2. 湖南大学 信息科学与工程学院, 湖南 长沙 410082)

摘要: 并行坐标下降(parallel coordinate descent, PCD)是一种最优秀的降噪算法,它被广泛地应用于图像处理中。由于算法有关参数的不确定性,使得 PCD 在音频信号降噪中并未获得相应的运用。为此,通过仿真实验,比较在音频环境下不同噪声强度、不同软阈值、不同超完备字典以及不同迭代次数对 PCD 降噪性能的影响,以提供算法相关参数的选择依据,从而扩展 PCD 的应用范围。实验结果表明,输入信噪比 $SNR=10$ (或加性 Gauss 白噪声的标准差 $\sigma=0.1$)是 PCD 对音频处理的极限条件。在全局阈值规则下,当收缩函数的阈值取 $\lambda=0.1$ 时,PCD 的降噪效果最佳。在冗余表示的基础上,采用 Gabor 超完备字典对信号分解可使 PCD 的降噪效果更好。当算法迭代次数到达 $M=25$ 时,PCD 获得稳定的收敛。

关键词: 冗余表示; 降噪; 并行坐标下降; 迭代收缩; 超完备字典

中图分类号: TN911.4, TN911.7 **文献标志码:** A **文章编号:** 1006-4710(2021)02-0215-07

Parameters selection of PCD algorithm for audio signal de-noising

HE Xuansen^{1,2}, XU Li¹, FAN Yueping¹

(1. School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou 511363, China;

2. College of Information Science and Engineering, Hunan University, Changsha 410082, China)

Abstract: The parallel coordinate descent (PCD) is one of the best de-noising algorithms, which is widely used in image processing. Due to the uncertainty of algorithm parameters, PCD has not been used in audio signal de-noising. Therefore, through simulation experiments, the effects of different noise intensities, different soft thresholds, different over-complete dictionaries, and different iterations on the de-noising performance of PCD in audio environment are compared so as to provide a basis for the selection of algorithm related parameters and expand the application scope of PCD. The experimental results show that the input signal-to-noise ratio $SNR=10$ (or the standard deviation of additive white Gaussian noise $\sigma=0.1$) is the limit condition of PCD for audio processing. Under the global threshold rule, when the threshold value of the shrinkage function is $\lambda=0.1$, the de-noising effect of PCD is the best. On the basis of redundant representation, using Gabor over-complete dictionary to decompose the signal can make the de-noising effect of PCD better. When the number of iterations reaches $M=25$, PCD algorithm can have stable convergence.

Key words: redundant representation; de-noising; parallel coordinate descent; iterative shrinkage; over-complete dictionary

降噪是信号处理最重要的任务之一,而对信号的分解有助于降噪。利用字典可以将信号分解成一系列原子的线性组合^[1]。以超完备字典^[2]为基础,冗余(稀疏)表示提供了一种强有力的信号模型^[3],它将信号近似为超完备字典中最少数量非零原子的

线性组合^[4]。信号的稀疏表示具有潜在的降噪能力^[5],以冗余表示为基础的收缩^[6]成为具有吸引力的降噪技术。小波收缩^[7]使用正交变换实现降噪,而新的趋势则是采用超完备变换。在众多信号分解方法中,基本追踪(basis pursuit, BP)^[8]法是将信号

收稿日期: 2020-04-01; 网络出版日期: 2020-10-26

网络出版地址: <https://kns.cnki.net/kcms/detail/61.1294.N.20201023.1745.006.html>

基金项目: 国家自然科学基金资助项目(60572183)

第一作者: 何选森, 男, 教授, 研究方向为统计信号处理与盲源分离。E-mail: 1410749772@qq.com

分解为字典原子的最佳叠加,并使表示系数的 l_1 范数(或 l_0 范数)最小化^[9]。对于消除 Gauss 白噪声,基本追踪降噪(BP denoising, BPDN)^[10]是最佳的。基于超完备字典和冗余表示,迭代收缩(iterative shrinkage, IS)算法^[11]成为一种高效的数值技术。文献[12]提出了四种 IS 算法,其中并行坐标下降(parallel coordinate descent, PCD)法具有最快的搜索速度。文献[13]对 PCD 进行了收敛性分析,并引入正则化函数的形式,使得对坐标优化提供解析的解。PCD 在图像处理^[14]和图像降噪^[15]方面获得了广泛的应用。

尽管 PCD 被证明是一种优秀的图像降噪算法,但它在音频降噪领域并没有得到相应的运用。这是因为在音频或语音信号降噪时,需要事先对 PCD 算法的参数进行设定,而目前还没有文献给出相关参数的设置标准。正是基于这种考虑,本文利用仿真实验,研究 PCD 算法相关参数的选择依据。主要通过比较音频信号外加 Gauss 噪声的强度、软收缩函数的阈值、算法的迭代次数、分解信号的超完备字典与 PCD 降噪性能之间的关系,给出以上参数选择的基本原则,以扩展 PCD 算法的应用范围。

1 去噪背景

设 $\mathbf{x} \in \mathbf{R}^N$ 为 N 维观测信号矢量, $\mathbf{s} \in \mathbf{R}^N$ 为信源, $\mathbf{T} \in \mathbf{R}^{N \times N}$ 为正交变换,则观测信号表示为 $\mathbf{x} = \mathbf{T}\mathbf{s}$ 。对于稀疏信源 \mathbf{s} ,在同一时刻,只有很少分量 $s_i \in \mathbf{s}$ 是非零值,其它多数分量都为零值。在实际应用中,信源是不可观测的,只能通过传感器采集来获得观测信号 \mathbf{x} 。由于采集信道噪声的存在, \mathbf{x} 中不可避免地包含了噪声。对于稀疏信号,小波分析中的收缩降噪方法^[16]是假定绝对值很小的分量为噪声并将其设置为零,而仅保留较大绝对值的分量以实现降噪目的。

若信号 $\mathbf{x} \in \mathbf{R}^N$ 可以由几个(单位能量的)原子 \mathbf{d}_k 的线性组合来近似,则由所有原子组成的集合 $\{\mathbf{d}_k : k=1, 2, \dots, K\}$ 就是一个字典 $\mathbf{D} \in \mathbf{R}^{N \times K}$ 。设矢量 $\mathbf{z} \in \mathbf{R}^K$ 是稀疏的,则 \mathbf{x} 可表示为:

$$\mathbf{x} = \mathbf{D}\mathbf{z} \quad (1)$$

式中: \mathbf{z} 是信号 \mathbf{x} 的重构系数。

当 $K=N$ 时,字典 \mathbf{D} 是正交的,则式(1)的表示是唯一的。当 $K>N$ 时,字典 \mathbf{D} 是超完备的,这种情况下式(1)存在有无限多个可能的系数集。这种非唯一性表示提供了极大的选择性,即从众多表示中选择最适合的表示。寻找具有最少量非零系数的解是一种非常具有吸引力的优化技术^[17]。当使用

超完备字典分解信号时,式(1)给出了欠定的线性方程组,它的最稀疏解可表示为如下的优化问题^[17]:

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s. t. } \mathbf{x} = \mathbf{D}\mathbf{z} \quad (2)$$

或者

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s. t. } \|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2 \leq \epsilon \quad (3)$$

式中: ϵ 是误差容限; $\|\cdot\|_p$ 为 l_p 范数; l_0 范数是对一个向量非零元素的计数。

式(2)~(3)中的优化问题可以通过正交匹配追踪法^[12]求解。基本追踪(BP)是寻找系数具有最小 l_1 范数的信号表示^[9],若在式(2)中用 l_1 范数代替 l_0 范数,则有:

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{s. t. } \mathbf{x} = \mathbf{D}\mathbf{z} \quad (4)$$

式中: l_1 范数是求向量各元素绝对值的和。

因 l_1 范数是 l_0 范数的凸化^[17],也就是说式(4)是式(2)的凸化。为了更接近于实际应用的环境,考虑一个含噪声的信号矢量

$$\mathbf{y} = \mathbf{x} + \mathbf{v} \quad (5)$$

式中: $\mathbf{y} \in \mathbf{R}^N$ 是信号 \mathbf{x} 的含噪声观测量; $\mathbf{v} \in \mathbf{R}^N$ 是 Gauss 零均值白噪声向量,并且限制它的能量为有限值。对信号 \mathbf{y} 降噪的优化问题表示为^[18-19]:

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{s. t. } \|\mathbf{y} - \mathbf{D}\mathbf{z}\|_2 \leq \delta \quad (6)$$

此问题可以由 Lagrange 形式替代^[9]:

$$\min_{\mathbf{z}} \frac{1}{2} \|\mathbf{D}\mathbf{z} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z}\|_1 \quad (7)$$

式中: λ 是一个正的参数,用于在近似误差与系数向量稀疏性二者之间寻求平衡。式(7)就是 BP-DN^[10],降噪后它的输出为 $\hat{\mathbf{x}} = \mathbf{D}\mathbf{z}_M$,其中 M 是预先设置的算法最大迭代次数, \mathbf{z}_M 是 \mathbf{z} 的最终迭代值,即式(7)的最稀疏解。

2 超完备字典

在 PCD 算法中,超完备字典起着非常重要的作用,选择合适的超完备字典是 PCD 实现收缩降噪的前提条件。

在音频信号处理领域,有两种典型的超完备字典,分别是离散余弦变换(discrete cosine transform, DCT)字典^[20]和 Gabor 字典^[20]。

一个窗口式 DCT 字典表示为 $\mathbf{D}^{\text{DCT}} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K]$,它的第 k 个原子定义为^[20]

$$\mathbf{d}_k(t) = \mathbf{w}_d(t) \cos \left[\frac{\pi}{K} \cdot \left(t + \frac{1}{2} \right) \cdot \left(k + \frac{1}{2} \right) \right] \quad (8)$$

式中: $1 \leq k \leq K$; $0 \leq t \leq N-1$; N 是每个音频帧的长度; K 是 DCT 字典的大小(尺寸),也就是离散频率的数量; \mathbf{w}_d 是加权窗口。

DCT 原子通过引入初始相位 φ 可推广为 Gabor 原子,这样做是为了增强信号结构的潜在适应性和信号的稀疏性。

窗口式 Gabor 字典表示为 $\mathbf{D}^G = [\mathbf{d}_{(k, \varphi)}]_{(k, \varphi) \in \Psi}$, 原子以连续参数集 $\Psi = [1, K] \times [0, 2\pi]$ 为索引, 定义为^[20]:

$$\mathbf{d}_{k, \varphi}(t) = \mathbf{w}_d(t) \cos \left[\frac{\pi}{K} \cdot \left(t + \frac{1}{2} \right) \cdot \left(k + \frac{1}{2} \right) + \varphi \right] \quad (9)$$

式中: $1 \leq k \leq K$; $0 \leq t \leq N-1$; N 是帧的长度; K 是 Gabor 字典的尺寸大小; φ 是它的相位; \mathbf{w}_d 是其加权窗口。在无任何近似的情况下, Gabor 字典连续索引的原子 $\mathbf{d}_{(k, \varphi)}$ 的分解可以用离散字典中的原子对来表示。一对的原子可以是相同频率且相位为零的余弦和正弦对:

$$\mathbf{d}_k^s(t) = \mathbf{w}_d(t) \sin \left[\frac{\pi}{K} \cdot \left(t + \frac{1}{2} \right) \cdot \left(k + \frac{1}{2} \right) \right] \quad (10)$$

$$\mathbf{d}_k^c(t) = \mathbf{w}_d(t) \cos \left[\frac{\pi}{K} \cdot \left(t + \frac{1}{2} \right) \cdot \left(k + \frac{1}{2} \right) \right] \quad (11)$$

一般地,对 DCT 和 Gabor 词典的列向量(音频帧)都进行归一化处理以简化其计算。例如式(10)~(11)所示 Gabor 字典的正弦和余弦原子,采用它们对应的单位范数版本。

3 PCD 算法

实际上,式(7)就是 PCD 算法的目标函数^[12]。对于含噪声信号 \mathbf{y} , 去除其中的 Gauss 噪声可归纳为对以下函数的最小化^[12]:

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \rho(\mathbf{x}) \quad (12)$$

当采用最大后验概率估计时,函数 $f(\mathbf{x})$ 可从 Bayes 的观点来解释。在式(12)中,等号右边第一项通常被称为对数似然,它描述干净信号 \mathbf{x} 与其含噪声观测量 \mathbf{y} 之间的关系,而第二项 $\rho(\mathbf{x})$ 表示未知信号 \mathbf{x} 的先验值。对不同的函数 $\rho(\mathbf{x})$, 式(12)中 $f(\mathbf{x})$ 具有不同的展开式。如果字典 \mathbf{D} 是超完备的, 则式(12)的目标函数将变成以下的优化问题^[12]:

$$\min_{\mathbf{z}} \frac{1}{2} \|\mathbf{D}\mathbf{z} - \mathbf{y}\|_2^2 + \lambda \rho(\mathbf{z}) \quad (13)$$

将优化式(13)与式(7)比较可知,在 PCD 算法中,函数 $\rho(\mathbf{z})$ 采用的是 l_1 范数。在文献^[12]中详细地介绍了 PCD 算法的推导过程,这里仅给出最终的更新迭代式:

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \mu \{S[\mathbf{v}(\mathbf{D}, \mathbf{y}, \mathbf{z}_k)] - \mathbf{z}_k\} = \mathbf{z}_k + \mu \mathbf{h}_k \quad (14)$$

式中: \mathbf{h}_k 是可计算的向量; \mathbf{z}_k 是 \mathbf{z} 的第 k 次解, \mathbf{z}_{k+1} 是其更新值; μ 是步长参数,若 $\mu = 1$, 则采用完全收缩,若 $\mu < 1$, 则效果为松弛步长; S 是软收缩(阈值)函数,其定义为:

$$S(z, \theta) = \begin{cases} z + \theta & (z < -\theta) \\ 0 & (|z| \leq \theta) \\ z - \theta & (z > \theta) \end{cases} \quad (15)$$

式中: θ 为阈值。

在式(14)中,向量 $\mathbf{v}(\mathbf{D}, \mathbf{y}, \mathbf{z}_k)$ 定义为:

$$\mathbf{v}(\mathbf{D}, \mathbf{y}, \mathbf{z}_k) = \text{diag}^{-1}(\mathbf{D}^T \mathbf{D}) \mathbf{D}^T (\mathbf{y} - \mathbf{D} \mathbf{z}_k) + \mathbf{z}_k \quad (16)$$

由式(14)可知收缩函数 S 以阈值 $\lambda \cdot \text{diag}^{-1}(\mathbf{D}^T \mathbf{D}) \cdot \mathbf{1}_K$ 对向量 $\mathbf{v}(\mathbf{D}, \mathbf{y}, \mathbf{z}_k)$ 逐项操作,其中 $\mathbf{1}_K$ 表示元素为 1 的 $K \times 1$ 矩阵(列向量)。由于字典 \mathbf{D} 的列是归一化的, 则 $\text{diag}^{-1}(\mathbf{D}^T \mathbf{D}) = \mathbf{I}$, 其中 \mathbf{I} 为单位矩阵。于是,式(16)可简化为:

$$\mathbf{v}(\mathbf{D}, \mathbf{y}, \mathbf{z}_k) = \mathbf{D}^T (\mathbf{y} - \mathbf{D} \mathbf{z}_k) + \mathbf{z}_k \quad (17)$$

PCD 算法的搜索速度可以通过线搜索^[11]或序列子空间优化^[19]来进一步加快。线搜索即一维搜索,是指单变量函数的优化过程。在 PCD 算法中,线搜索用于寻找最佳的步长 μ , 其目的是求解以下的优化问题:

$$\min_{\mu} \frac{1}{2} \|\mathbf{D}(\mathbf{z}_k + \mu \mathbf{h}_k) - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z}_k + \mu \mathbf{h}_k\|_1 \quad (18)$$

而式(18)的解是通过求下面的方程获得:

$$\mathbf{h}_k^T \mathbf{D}^T [\mathbf{D}(\mathbf{z}_k + \mu \mathbf{h}_k) - \mathbf{y}] + \lambda \mathbf{h}_k^T \cdot \text{sign}(\mathbf{z}_k + \mu \mathbf{h}_k) = 0 \quad (19)$$

式中: sign 为符号函数。式(15)也具有收缩的结构。

在 PCD 算法中, \mathbf{z}_k 的初始值为 $\mathbf{z}_0 = \mathbf{0}_K$, 其中 $\mathbf{0}_K$ 表示元素为 0 的 $K \times 1$ 矩阵, 经过 M 次迭代后输出为 $\hat{\mathbf{x}} = \mathbf{D}\mathbf{z}_M$, 即经过降噪后获得的干净信号。

4 仿真结果与分析

4.1 仿真环境与性能指标

为了验证 PCD 算法对音频信号的降噪性能,通过一系列仿真实验,归纳出算法相关参数的选择原则。仿真 PC 机配置为: Intel (R) Celeron (R) 1007U-1.5 GHz 的 CPU, 4 GB 内存, Windows 10 操作系统。所有仿真都是在 MATLAB 9 (R2016a) 上运行。用于测试的音频源信号为两通道的音乐信号,其类型为 wav, 信号的采样率为 44.1 MHz。每个信源样本数为 $L = 2^{16} = 65\ 536$ 。五个音频信源

$\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5]^T$ 的时域波形见图1。

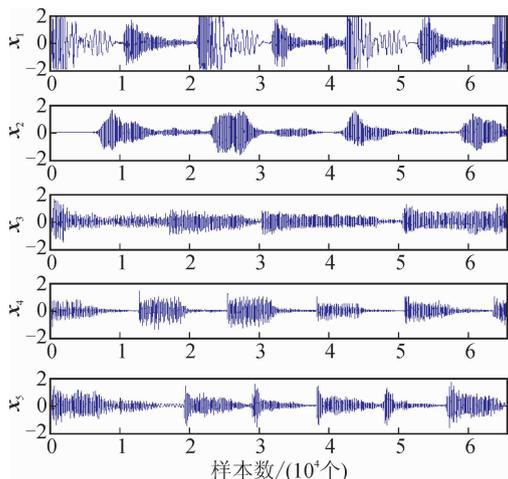


图1 五路音乐音频源信号的波形

Fig. 1 The time domain waveform of 5 source signals

为了测度PCD算法的降噪性能,采用两种经典的技术指标,分别是信噪比(signal to noise ratio, SNR)和均方根误差(root mean square error, RMSE):

$$SNR = 10 \log_{10} \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \quad (20)$$

$$RMSE = \sqrt{\frac{1}{L} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \quad (21)$$

式中: \mathbf{x} 是原始的无噪信号; $\hat{\mathbf{x}}$ 是降噪后信号; L 是 \mathbf{x} 的样本数量。从式(20)可知,SNR指标给出的是分贝(dB)数。

信号降噪的仿真过程如下。在时域中音频信源与Gauss噪声相加生成含噪声的观测信号。每个观测信号乘以长度为64的Hamming窗以生成音频(语音)帧,使连续两帧之间的重叠样本数为32,即每一帧的前后各有16个样本与相邻帧的样本重叠。音频帧的分割过程是通过使用Voicebox中的MATLAB函数enframe来实现。将PCD算法应用于分帧后的观测信号进行降噪,每次使用PCD是对一帧进行处理。对每个降噪的帧乘以逆Hamming窗,且在相邻帧重叠的样本中,分别在每一帧两边各去除50%的样本,仅保留每一帧中间部分的样本,最后将经过处理的各个片段简单串联起来即可得到重新合成的降噪后的信号。

PCD算法的迭代过程如下。设定最大迭代次数 M 。将迭代次数初始化为 $k=1$,仿真开始。在PCD算法每次迭代后计算并记录相应的RMSE和SNR值。然后,将迭代次数更新为 $k \leftarrow k+1$,以重复执行PCD算法。最后,判断仿真过程是否结束,若 $k < M$,则仿真继续;若 $k = M$,则仿真结束,并给出降噪后的结果。

4.2 噪声强度对降噪性能的影响

在PCD算法的迭代中,需要预先给出软收缩函数 S 的阈值 λ 。在确定最优的阈值 λ 时,可以使用全局阈值规则和最小最大阈值规则^[7]。本文采用全局阈值规则,其 λ 的取值范围为^[7]:

$$\lambda \leq \sigma \sqrt{2 \log_2 L} \quad (22)$$

式中: L 是信号的样本长度; σ 是Gauss噪声的标准差。从式(22)可知,要确定阈值 λ ,就必须先要给出噪声的强度 σ 。

音频信源在被加入Gauss噪声后生成观测信号,而加入的噪声强度采用(input)信噪比来度量。为确定噪声强度对降噪性能的影响,本小节采用小波软阈值降噪法。小波的类型为db45,当观测信号的输入(input)SNR=10, 15, 20dB时,利用小波降噪获得相应的输出SNR值和RMSE值。在这个仿真中,对于五个信源的加噪、降噪过程重复进行了100次实验。在每次实验后,分别计算并记录全部五个观测信号在一次实验中平均的输出SNR和RMSE值。仿真实验的结果见图2。

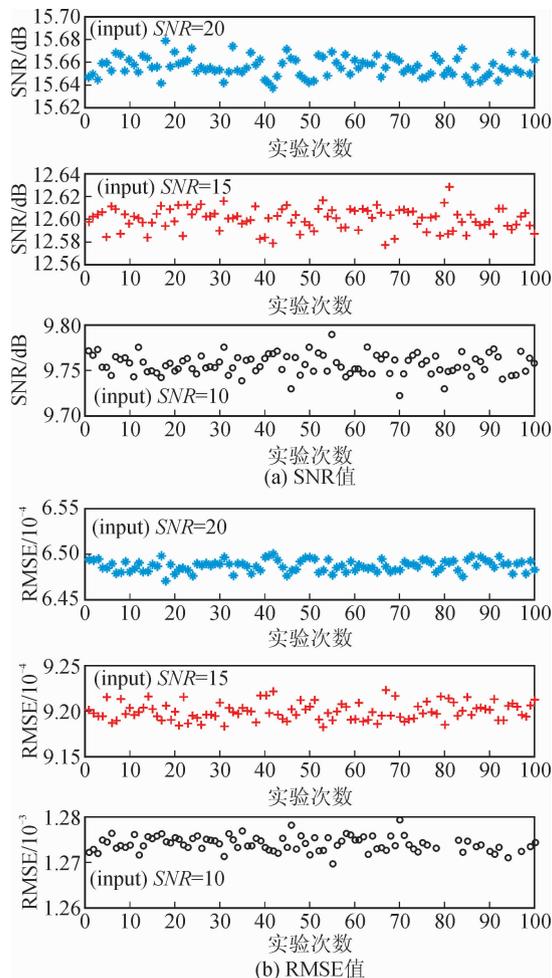


图2 五个信号的平均输出SNR和RMSE值

Fig. 2 Average output SNR and RMSE of five signals

从图2的平均性能指标可以看出,随着观测信号中SNR值的降低(即噪声强度增大),小波软阈值方法的降噪性能变差。特别地,当(input)SNR=10时,降噪后输出的平均SNR<9.8,即经过降噪处理的信噪比小于降噪前的信噪比。这说明由于噪声强度过大,信号被严重污染而不能恢复。对音频信号来说,(input)SNR=10是一个非常恶劣的环境。

为便于分析和比较,对每一个输入的SNR值,分别计算施加到每个信源上的噪声强度 σ 。其对应的结果见表1。

表1 输入SNR与噪声强度 σ 的对应关系

Tab.1 Correspondence between input SNR and noise intensity

(input) SNR	每个信号的噪声强度 σ				
	x_1	x_2	x_3	x_4	x_5
10	0.200 1	0.099 5	0.090 0	0.070 0	0.082 6
15	0.112 5	0.055 9	0.050 6	0.039 4	0.046 4
20	0.063 3	0.031 5	0.028 5	0.022 1	0.026 1

从表1的数据可以看出,对于相同的输入SNR值,不同信源被施加的噪声强度是不相同的,这是因为每个信源的幅度(或功率)都不相同所致。对于(input)SNR=10的最坏情况,施加到信号 x_2, x_3, x_4, x_5 上的噪声强度接近0.1,而施加到信号 x_1 的噪声强度为0.2。为统一起见,在下面的仿真实验中,将相同的噪声强度($\sigma=0.1$)施加于五个信源形成含噪声的观测信号,也就是PCD算法的输入信号。

4.3 软阈值对降噪性能的影响

在给定噪声强度 σ 之后,利用式(22)就可以求出软收缩函数S的阈值 λ 。

因为信号的采样点数为: $L=2^{16}=65\ 536$,因此有 $(2\log_2 L)^{1/2}=4.7096$,从文献[7]中的Table 2可以知道,在给定 $\sigma=1$ 时,其参考的阈值为 $\lambda^*=3.310$ 。由于本文中选取 $\sigma=0.1$,将它代入到式(22)可得结果: $\lambda \leq \sigma \times 3.310=0.3310$,显然这只是 λ 的一个取值范围。为了找出PCD算法的最佳阈值 λ ,通过以下的仿真实验来进行确定。其实验条件为:每个观测信号都被分割成长度为64的音频帧,设置PCD算法的最大迭代次数为 $M=5$ 。软阈值 λ 的值从0.05到0.3310不断地变化,其更新步长为0.01。超完备字典分别采用DCT和Gabor,对每个 λ 值和每个字典仅进行一次实验。图3给出了在这种条件下PCD算法输出的SNR和RMSE值。

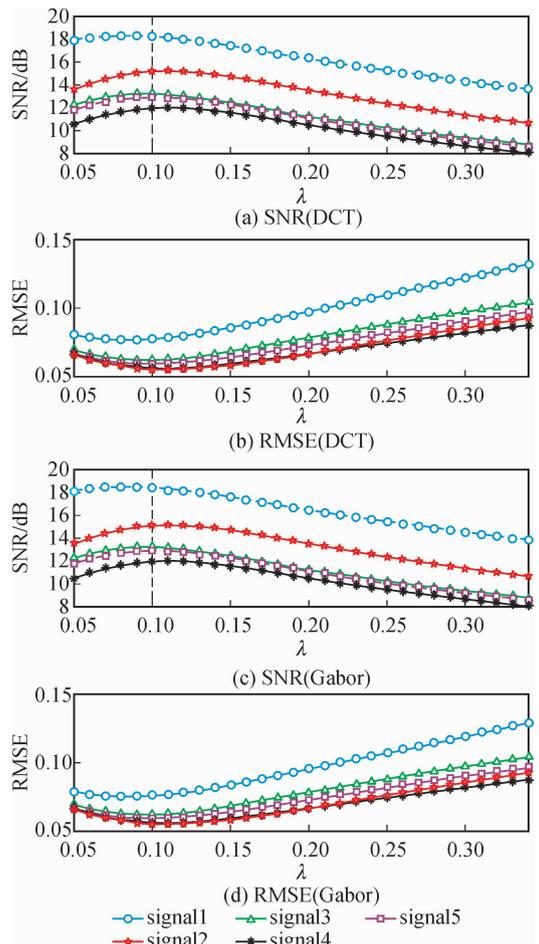


图3 五个信号对不同 λ 值的输出SNR和RMSE
Fig.3 Output SNR and RMSE of five signals for different λ 's

从图3的结果可以看出,无论采用DCT还是Gabor字典,PCD算法对每个观测信号的降噪性能,即输出的SNR和RMSE值随着阈值 λ 的增加都有明确的变化趋势。显然地,当 $\lambda=0.1$ 时每个信号的输出SNR达到相应的最大值,而每个信号的输出RMSE也达到相对应的最小值。这说明当选取阈值 $\lambda=0.1$ 时,PCD算法能获得最好的降噪效果。

4.4 超完备字典对降噪性能的影响

从图3还可看出,对于不同的超完备字典,PCD算法的去噪性能是不相同的。因此,在给定噪声强度 σ 和软收缩函数的阈值 λ 之后,还需要分析超完备字典对PCD算法降噪性能的影响,以确定最适合于PCD算法的字典。

在这个仿真中,PCD算法的最大迭代次数设置为 $M=15$ 。分别采用DCT和Gabor字典,对每个观测信号利用PCD算法进行降噪处理,并计算和记录每个信号在每次迭代中的性能指标。图4给出了实验的结果(输出的SNR和RMSE值)。

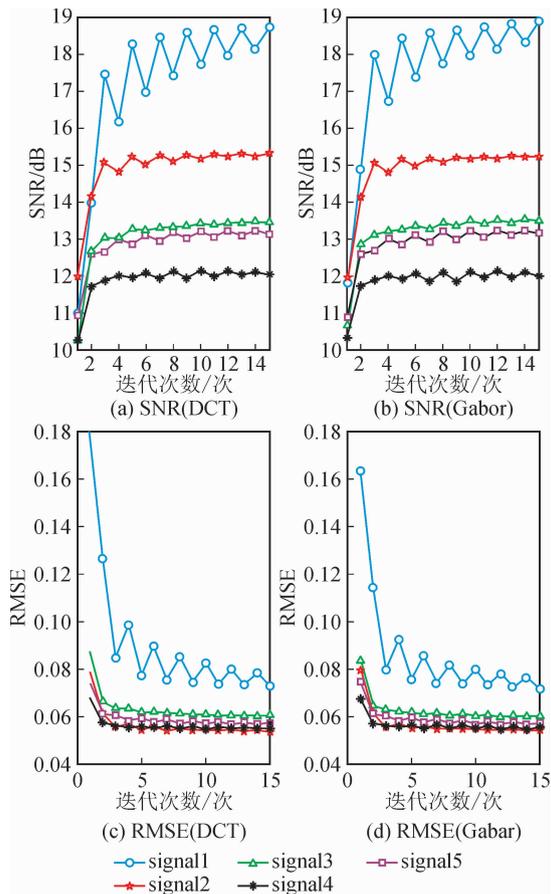


图 4 五个信号对不同字典的输出 SNR 和 RMSE
Fig. 4 Output SNR and RMSE of five signals for different dictionaries

从图 4 的输出指标波形可以看出,随着迭代次数的增加,PCD 算法的降噪性能逐渐得到提高,即 PCD 算法随着迭代次数的增加而逐渐收敛。观察第一个信号(即 signal1)的波形可知,使用 Gabor 字典所得到的 SNR 值要大于使用 DCT 字典对应的 SNR 值。同样地,使用 Gabor 字典得到的 RMSE 值要小于使用 DCT 字典的 RMSE 值。此外,信号 2 至信号 5 输出的 SNR 和 RMSE 值也都具有相类似的关系。这意味着使用 Gabor 字典,PCD 算法可以获得更好的降噪性能。

4.5 算法迭代次数对降噪性能的影响

在给定噪声强度 σ ,软阈值 λ ,并采用 Gabor 超完备字典的情况下,研究 PCD 算法的降噪性能与算法迭代次数的关系。在这个仿真中,PCD 算法的最大迭代次数设为 $M=50$,Gabor 字典的尺寸为 $N \times K = 64 \times 1024$,它的加权窗口 w_d 为长度 $N=64$ 的对称正弦窗口。在 PCD 算法的每次迭代过程中,分别记录每个信号的输出 SNR 和 RMSE 值,其结果见图 5。

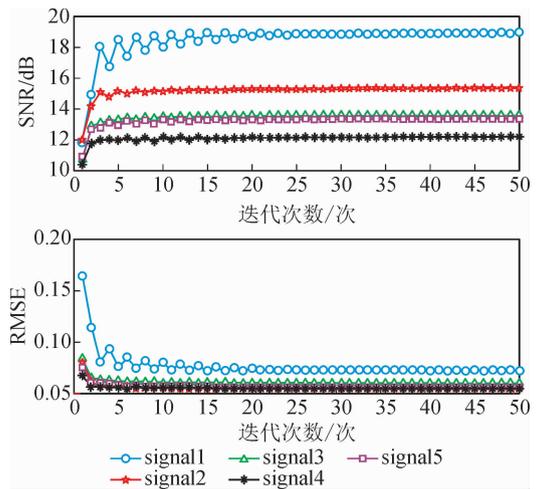


图 5 五个信号在每次迭代后的输出 SNR 和 RMSE
Fig. 5 Output SNR and RMSE of five signals after every iteration

从图 5 的输出结果可以看出,对于每一个观测信号,在经过大约 25 次迭代后,PCD 算法降噪的指标 SNR 和 RMSE 值达到相对应的稳定值。也就是说,以降噪为目标的 PCD 算法,其收敛条件为算法需要经过 25 次的优化迭代。

5 结 语

针对音频信号的降噪问题,本文通过仿真,比较分析了算法的相关参数对 PCD 降噪性能的影响。仿真结果表明如下结论。

- 1) Gauss 白噪声的标准差为 0.1(或输入的 $SNR=10$)时会造成小波降噪性能的严重下降,这是因为音频信号本身的幅度(功率)较小所致,此结论对音频信号处理具有指导性意义。
- 2) 当软收缩函数的阈值为 0.1、采用 Gabor 字典时 PCD 算法能取得很好的降噪性能。
- 3) 当迭代次数为 25 时,PCD 算法能获得稳定的收敛。

参考文献:

- [1] YANG Jicheng, HE Qianhua, LI Yanxiong, et al. Dictionary learning based on M-PCA-N for audio signal sparse representation[J]. IET Signal Processing, 2018, 12(2): 198-206.
- [2] AGARWAL A, ANANDKUMAR A, NETRAPALLI P. A clustering approach to learning sparsely used overcomplete dictionaries[J]. IEEE Transactions on Information Theory, 2017, 63(1): 575-592.
- [3] MURATA N, KOYAMA S, TAKAMUNE N, et al. Sparse representation using multidimensional mixed-norm penalty with application to sound field decomposi-

- tion [J]. IEEE Transactions on Signal Processing, 2018, 66(12): 3327-3338.
- [4] WANG Gaoang, HWANG J N, ROSE C, et al. Uncertainty-based active learning via sparse modeling for image classification [J]. IEEE Transactions on Image Processing, 2019, 28(1): 316-329.
- [5] RAPHAN M, SIMONCELLI E P. Optimal denoising in redundant representations [J]. IEEE Transactions on Image Processing, 2008, 17(8): 1342-1352.
- [6] ISOGAWA K, IDA T, SHIODERA T, et al. Deep shrinkage convolutional neural network for adaptive noise reduction [J]. IEEE Signal Processing Letters, 2018, 25(2): 224-228.
- [7] DONOHO D L, JOHNSTONE I M. Ideal spatial adaptation via wavelet shrinkage [J]. Biometrika, 1994, 81(3): 425-455.
- [8] GOLDSTEIN T, STUDER C. PhaseMax: convex phase retrieval via basis pursuit [J]. IEEE Transactions on Information Theory, 2018, 64(4): 2675-2689.
- [9] ZIBULEVSKY M, ELAD M. L_1 - L_2 optimization in signal and image processing [J]. IEEE Signal Processing Magazine, 2010, 27(3): 76-88.
- [10] MOSHTAGHPOUR A, JACQUES L, CAMBARERI V, et al. Consistent basis pursuit for signal and matrix estimates in quantized compressed sensing [J]. IEEE Signal Processing Letters, 2016, 23(1): 25-29.
- [11] ZIBETTI M V W, HELOU E S, PIPA D R. Accelerating overrelaxed and monotone fast iterative shrinkage-thresholding algorithms with line search for sparse reconstructions [J]. IEEE Transactions on Image Processing, 2017, 26(7): 3569-3578.
- [12] ELAD M. Why simple shrinkage is still relevant for redundant representations? [J]. IEEE Transactions on Information Theory, 2006, 52(12): 5559-5569.
- [13] ELAD M, MATALON B, ZIBULEVSKY M. Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization [J]. Applied and Computational Harmonic Analysis, 2007, 23(3): 346-367.
- [14] LIU Baodi, WANG Yuxiong, SHEN Bin, et al. Blockwise coordinate descent schemes for efficient and effective dictionary learning [J]. Neurocomputing, 2016, 178: 25-35.
- [15] LI Zhenni, DING Shuxue, HAYASHI T, et al. Analysis dictionary learning using block coordinate descent framework with proximal operators [J]. Neurocomputing, 2017, 239: 165-180.
- [16] DONOHO D L, JOHNSTONE I M. Adapting to unknown smoothness via wavelet shrinkage [J]. Journal of the American Statistical Association, 1995, 90(432): 1200-1224.
- [17] BABAIE-ZADEH M, JUTTEN C. On the stable recovery of the sparsest overcomplete representations in presence of noise [J]. IEEE Transactions on Signal Processing, 2010, 58(10): 5396-5400.
- [18] METZLER C A, MALEKI A, BARANIUK R G. From denoising to compressed sensing [J]. IEEE Transactions on Information Theory, 2016, 62(9): 5117-5144.
- [19] LEI Gang, GUO Y G, ZHU J G., et al. Sequential subspace optimization method for electromagnetic devices design with orthogonal design technique [J]. IEEE Transactions on Magnetics, 2012, 48(2): 479-482.
- [20] ADLER A, EMIYA V, JAFARI M G, et al. Audio inpainting [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(3): 922-932.

(责任编辑 王绪迪)