

带边界条件约束的非相干字典学习方法及其稀疏表示

汤红忠^{1,2,4} 张小刚¹ 陈华³ 程炜^{2,4} 唐美玲¹

摘要 从字典的相干性边界条件出发, 提出一种基于极分解的非相干字典学习方法 (Polar decomposition based incoherent dictionary learning, PDIDL), 该方法将字典以 Frobenius 范数逼近由矩阵极分解获取的紧框架, 同时采用最小化所有原子对的内积平方和作为约束, 以降低字典的相干性, 并保持更新前后字典结构的整体相似特性. 采用最速梯度下降法和子空间旋转实现非相干字典的学习和优化. 最后将该方法应用于合成数据与实际语音数据的稀疏表示. 实验结果表明, 本文方法学习的字典能逼近等角紧框架 (Equiangular tight-frame, ETF), 实现最大化稀疏编码, 在降低字典相干性的同时具有较低的稀疏表示误差.

关键词 字典学习, 非相干字典, 等角紧框架, 稀疏表示

引用格式 汤红忠, 张小刚, 陈华, 程炜, 唐美玲. 带边界条件约束的非相干字典学习方法及其稀疏表示. 自动化学报, 2015, 41(2): 312–319

DOI 10.16383/j.aas.2015.c140183

Incoherent Dictionary Learning Method with Border Condition
Constrained for Sparse RepresentationTANG Hong-Zhong^{1,2,4} ZHANG Xiao-Gang¹ CHEN Hua³ CHENG Wei^{2,4} TANG Mei-Ling¹

Abstract A novel incoherent dictionary learning method is introduced based on polar decomposition, which is simple but practical framework for designing incoherent dictionary. The method distinguishes itself from traditional dictionary learning approaches by explicitly taking into account the border condition. In particular, the dictionary approximates the r -tight frame obtained with matrix polar decomposition with respect to Frobenius norm. The minimization of the sum of squared inner product of all atom pairs is taken as an constraint so as to reduce the dictionary coherence and keep the overall dictionary structure similarity before and after the update. Incoherent dictionary learning and optimization are performed based on the steepest gradient descent method and subspace rotation. Experiments on synthetic data and real audio data show that the proposed approach can achieve considerable improvements in terms of approximating equiangular tight frame (ETF) and implementing the maximal sparse coding, and can effective control the trade-off between low sparse representation errors and dictionary coherence in comparison to other methods.

Key words Dictionary learning, incoherent dictionary, equiangular tight-frame (ETF), sparse representation

Citation Tang Hong-Zhong, Zhang Xiao-Gang, Chen Hua, Cheng Wei, Tang Mei-Ling. Incoherent dictionary learning method with border condition constrained for sparse representation. *Acta Automatica Sinica*, 2015, 41(2): 312–319

近几年来, 字典学习的理论与应用研究受到极大关注. 它可提取样本数据的内部结构特征, 已被广

泛应用于去噪^[1]、特征提取^[2]、压缩感知^[3–4]、模式识别与分类^[5–6]等领域. 其学习方法一般可通过优化下式获取:

$$\arg \min_{D, X} \frac{1}{2} \|Y - DX\|_F^2$$

$$\text{s. t. } \|\mathbf{x}_i\|_0 \leq S, \quad i = 1, 2, \dots, K \quad (1)$$

其中, $Y \in \mathbf{R}^{M \times N}$ 为信号的样本数据矩阵, $Y = \{\mathbf{y}_i\}_{i=1}^N$, M, N 分别为样本数据的维数和个数; 矩阵 $D \in \mathbf{R}^{M \times K}$ 为过完备字典^[7] ($M < K$), $D = \{\mathbf{d}_k\}_{k=1}^K$, 其列 $\mathbf{d}_k$ 称为字典的原子; $X \in \mathbf{R}^{K \times N}$ 为信号的稀疏表示系数矩阵, $X = \{\mathbf{x}_i\}_{i=1}^N$, S 为 $\mathbf{x}_i$ 中非 0 元素的个数.

在上述字典学习过程中, 字典中各原子间的相干性是稀疏信号恢复的重要条件. Donoho 等^[8]指

收稿日期 2014-03-31 录用日期 2014-07-23

Manuscript received March 31, 2014; accepted July 23, 2014

国家自然科学基金 (61174050, 61203016), 湘潭大学控制科学与工程学科建设经费资助

Supported by National Natural Science Foundation of China (61174050, 61203016) and Control Science and Engineering Disciplinary Construction Funds of Xiangtan University

本文责任编辑 刘成林

Recommended by Associate Editor LIU Cheng-Lin

1. 湖南大学电气与信息工程学院 长沙 410082 2. 湘潭大学信息工程学院 湘潭 411105 3. 湖南大学信息科学与工程学院 长沙 410082 4. 湘潭大学控制工程研究所 湘潭 411105

1. College of Electrical and Information Engineering, Hunan University, Changsha 410082 2. College of Information Engineering, Xiangtan University, Xiangtan 411105 3. College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082 4. Institute of Control Engineering, Xiangtan University, Xiangtan 411105

出字典的相干性若满足一定的条件, 式 (1) 中的 l_0 范数可以等价用 l_1 范数求解. Tropp^[9] 在文献 [8] 的基础上, 进一步给出稀疏信号精确恢复的充分条件 (见式 (3)), 并提出在采用 l_1 范数代替 l_0 范数及 OMP (Orthogonal matching pursuit) 算法时, 若字典具有较低相干性, 稀疏信号恢复则具有更佳的支撑集与更快的残差衰减速度. Donoho 等^[10] 证明了利用 l_1 范数与 OMP 算法解决 l_0 范数的优化问题时两者之间的一致性, 亦指出若字典 D 的相干性较低时, LARS (Least angle regression stagewise) 也是优化 l_0 问题的有效方法之一.

但是, 目前主流的字典学习方法, 如 K-SVD (K-singular value decomposition, K-SVD)^[11] 无法有效降低字典的相干性. 该方法对于相干性超过阈值 μ_{th} 的原子, 直接由样本中其他数据代替, 这不仅增加了稀疏表示误差, 而且直接采用阈值控制无法降低字典相干性.

为解决这一问题, 诸多学者提出一些降低字典相干性的学习方法^[12-16]. Mailhe 等^[12] 提出基于矩阵邻近最小化字典相干性的字典学习方法 (Incoherent K-SVD, INK-SVD), 该方法首先对高相干性的原子实现聚类, 然后将相干性高于阈值的原子成对解耦、反复迭代, 直至字典的相干性满足其目标要求. 但在该方法中, 字典相干性的阈值取值越小, 解耦步骤增加越多, 计算复杂度越高. Barchiesi 等^[13] 提出基于迭代投影与旋转的非相干字典学习方法 (Iterative projections and rotations, IPR), 该文首先采用交替投影方法优化字典 D 的 Gram 矩阵, 构造 Grassmannian 框架, 获取低相干性的字典, 然后利用文献 [14-15] 方法对字典进行旋转优化, 减小稀疏表示误差. 由于这种方法需事先确定交替投影算法中每次投影时, Gram 矩阵中要更新的元素个数以及收缩因子的大小, 参数较难寻优. Cristian^[16] 提出一种非相干性框架的设计方法 (Iterative decorrelation by convex optimization, IDCO), 基本思想是直接优化字典 D 中原子间的最大内积, 该方法利用凸优化方法对相干性高的原子迭代解耦, 从而最小化字典的相干性. 但是, 原子间的最大内积, 仅能表示原子间的局部相似性, 因而这种方法在降低某原子相干性的同时, 可能会增加与其他原子间的相干性, 且字典更新时对单原子逐个更新, 随着字典维数的增加字典学习速度下降很快.

本文在文献 [16] 的基础上, 从字典的相干性边界条件出发, 利用矩阵极分解获取的 r 紧框架 Φ , 同时采用最小化所有原子对的内积平方和作为约束, 构造等角紧框架 (Equiangular tight-frame, ETF). 所有原子同时更新, 尽量使字典的原子结构符合信号本身所固有的特征; 同时利用正交矩阵与字典相

乘并不改变字典相干性的特点, 基于子空间旋转进一步优化字典. 将训练的非相干字典应用于信号的稀疏表示, 实验结果表明, 本文方法能有效降低字典的相干性, 同时具有较好的稀疏表示性能.

1 非相干字典

字典的相干性描述了过完备字典中原子间的最大相似程度, 定义为两个不同原子归一化后的内积的最大绝对值, 其数学公式如下^[8]:

$$\mu(D) = \max |\langle \mathbf{d}_i, \mathbf{d}_j \rangle|, \quad i \neq j \quad (2)$$

其中, $\mathbf{d}_i, \mathbf{d}_j$, 分别表示字典中两个不同原子. 同时, 字典 D 的 Gram 矩阵 $G = D^T D$ 中非对角线上的元素也表示原子间的内积, 因此, 相干性也等于 G 中非对角线上元素的最大绝对值. $\mu(D) \in [0, 1]$, $\mu(D)$ 的大小在一定程度上能反映原子间相似程度的强弱. 当 $\mu(D)$ 值很大, 表示原子间相似程度很强, 反之, 当 $\mu(D)$ 值很小, 表示原子间相似程度很弱, 此时, 将 D 称之为非相干字典. 尽管信号的稀疏性要求字典保持某种程度的相似性, 但这种相似性并不是越大越好, 稀疏表示中更受关注的是非相干字典.

在式 (1) 中, 已知 D , 如何求解稀疏系数 X 是一个 NP 难题, 解决这一问题有许多方法^[7, 17-18]. 在这些方法中, 字典 D 满足何种条件下, 其解更能逼近 X 的真实解成为研究重点之一. 有众多学者提供了一个稀疏信号精确恢复的充分条件^[8-10]:

定理 1. 给定字典 $D \in \mathbf{R}^{M \times K}$, 且 $M < K$, 其相干性为 $\mu(D)$, 假设 $\mathbf{y} = D\mathbf{x}$ 有稀疏解 $\mathbf{x}$, 其稀疏度 S 若满足

$$S < \frac{1}{2} \left(1 + \frac{1}{\mu(D)} \right) \quad (3)$$

则可推导出以下结论:

- 1) 解 $\mathbf{x}$ 必定是最稀疏的;
- 2) 式 (1) 中的 l_0 问题也可以等价于 l_1 问题求解;
- 3) OMP 能从字典 D 中找出最佳的 S 项原子的线性组合.

可以证明, 对于一般过完备字典的相干性存在一定的范围, 下界为 Welch Bound^[19], 其定理描述如下:

定理 2. 假设字典 $D \in \mathbf{R}^{M \times K}$ 且 $M < K$, 列向量归一化, 则字典 D 的相干性满足:

$$\mu(D) \geq \sqrt{\frac{K-M}{M(K-1)}} \quad (4)$$

若上式边界条件成立, 字典 D 的相干性达到最小, 越接近于正交基. 因此, 如何将字典的相干性逼

近其边界值成为稀疏信号恢复与压缩感知问题的关键.

Tropp 等^[20] 指出当且仅当 D 是等角紧框架, 且满足 $K \leq M(M+1)/2$, 式 (4) 中的等号成立. ETF 的定义如下:

定义 1. 假设 $D \in \mathbf{R}^{M \times K}$ 且 $M < K$, $D = \{\mathbf{d}_k\}_{k=1}^K$, 列向量归一化, 若满足:

$$|\mathbf{d}_i^T \mathbf{d}_j| = \mu, \quad i \neq j \quad (5)$$

且

$$DD^T = \frac{K}{M}I \quad (6)$$

则 D 被称为等角紧框架.

由式 (6) 可知, 若 D 为 ETF, 则 D 为行满秩矩阵, 具有 M 个非 0 的奇异值, 其值都等于 $\sqrt{K/M}$.

由式 (3) 与式 (4) 可知, 如果给定信号维数 256, 原子个数 512, 则 $\mu(D) > 0.04$, 则保证信号精确恢复的稀疏度满足 $S \leq 13$, 若 $\mu(D)$ 达到最低边界, 表示字典 D 中原子间相似程度越弱, 字典越接近于正交, 稀疏表示时越能最大化其稀疏编码.

字典 D 的相干性不仅影响着稀疏编码的支撑集, 而且, 随着算法迭代次数的增加, 具有较小相干性的字典 D , 对应稀疏编码的残差范数衰减速度更快, 同时 $\mu(D)$ 限定了残差范数收敛的上边界^[9-10]. 然而, 在字典学习过程中, 每一次迭代时, 若仅仅引入降低字典相干性的约束条件, 会直接改变已训练好的适合于表示样本数据的字典 D 的内部结构, 虽然降低了相干性, 但字典的稀疏表示误差会增加. 因此, 如何有效训练相干性低 (泛化能力强) 且稀疏表示能力强的字典在信号分析感知领域有着潜在意义, 也是目前字典学习亟需解决的科学问题之一, 下面将给出本文提出的一种新的基于极分解的非相干字典学习方法.

2 本文方法

字典学习的主要优点在于给定样本数据, 通过训练过完备字典 D , 可同时得到适应信号的基与相应的稀疏表示系数. 式 (1) 目标函数中的 (D, X) 是一个组合的非凸优化问题, 其求解一般通过反复执行两步迭代来完成. 给定初始字典, 交替以下两个步骤直至满足收敛条件:

- 1) 稀疏编码: 固定字典 D , 求解 $\arg \min_X \frac{1}{2} \|Y - DX\|_F^2$ s.t. $\|\mathbf{x}_i\|_0 \leq S$, 计算系数 X ;
- 2) 字典更新: 固定系数 X , 求解 $\arg \min_D \frac{1}{2} \|Y - DX\|_F^2$ 更新字典 D .

2.1 稀疏编码

假设 D 固定, 更新 X , 本文利用 SPAMS 工具箱¹, LARS^[18] 方法求解稀疏表示系数. 目标函数梯度定义如下:

$$D^T(D\mathbf{x} - \mathbf{y}) + \eta \mathbf{u} \quad \forall \mathbf{u} = \begin{cases} 1, & \mathbf{x}[i] > 0 \\ [-1 \ 1], & \mathbf{x}[i] = 0 \\ -1, & \mathbf{x}[i] < 0 \end{cases} \quad (7)$$

其中, $\mathbf{y}$ 是 Y 中对应的某个样本数据, $\mathbf{x}[i]$ 是稀疏表示系数 $\mathbf{x}$ 中第 i 个元素, 则 $\mathbf{x}$ 的稀疏解为

$$\mathbf{x}_\eta^S = (D_S^T D_S)^{-1} (D_S^T \mathbf{y} - \eta \mathbf{u}_\eta^S) \quad (8)$$

S 是当前 η 值对应稀疏表示系数 $\mathbf{x}$ 的支撑集.

2.2 字典更新

由定义 1 与定理 2 可知, 符合 ETF 的字典能达到相干性的边界值, 但利用定义 1 直接构造 D 的等角紧框架比较困难. 而利用文献 [13] 的 Grassmannian 框架, 每次投影时都要确定 Gram 矩阵中要更新的元素个数以及收缩因子的大小, 参数较难寻优. 因此本文利用由矩阵极分解获取的 r 紧框架 Φ 来构造 D 的等角紧框架, 从而实现非相干字典的学习.

矩阵的极分解, 即字典 D 可分解为两个矩阵的乘积, 其定理如下^[21]:

定理 3. 假设字典 $D \in \mathbf{R}^{M \times K}$ 且 $M < K$, 存在 P 与 Q , 使得 $D = PQ$. 其中, $P \in \mathbf{R}^{M \times M}$, 是半正定阵, $\text{rank}(P) = \text{rank}(D)$, $Q \in \mathbf{R}^{M \times K}$, 具有正交行向量 ($QQ^T = I$).

证明.

根据矩阵的奇异值分解, $D = USV^T$, 其中, S 为对角阵, U, V 为正交阵. 则 $D = USU^T UV^T$, 令

$$P = USU^T, \quad Q = UV^T \quad (9)$$

可见, P 为半正定阵, $QQ^T = I$, 矩阵 Q 被称为字典 D 的正交极因子. $\square$

本文利用正交极因子 Q 构造了字典 D 的 r 紧框架, Tropp 给出了相应的定理^[20]:

定理 4. 假设 $D = PQ$, 且定义字典冗余度 $r = \frac{K}{M}$, 令 $\Phi = \sqrt{r}Q$, 则 Φ 是关于 Frobenius 范数的距离矩阵 D 最近的 r 紧框架.

可以证明, 基于定理 4 构造的 r 紧框架 Φ 与字典 D 的相干性边界条件相同.

证明.

¹<http://spams-devel.gforge.inria.fr/>

$$\begin{aligned}
0 &= \text{tr} \left[\left(\Phi \Phi^T - \frac{K}{M} \right)^2 \right] = \\
&\text{tr} \left[(\Phi \Phi^T)^2 - \frac{K^2}{M} \right] = \\
&\sum_i \sum_{i \neq j} |\langle \phi_i, \phi_j \rangle|^2 - \frac{K^2}{M} \leq \\
&K + K(K-1)\mu(\Phi)^2 - \frac{K^2}{M}
\end{aligned} \quad (10)$$

即

$$\mu(\Phi) \geq \sqrt{\frac{K-M}{M(K-1)}} \quad (11)$$

□

式 (11) 与式 (4) 相符, 说明矩阵 Φ 与字典 D 的相干性边界条件完全相同, 当且仅当矩阵 Φ 为 ETF 时, 上式中等号成立, 具有最小相干性。

基于上述讨论, 本文从式 (3) 的边界条件出发, 根据定理 4, 利用由矩阵极分解获取的 r 紧框架 Φ 来构造 D 的等角紧框架, 提出一种新的基于极分解的非相干字典学习方法 (Polar decomposition based incoherent dictionary learning, PDIDL), 其目标函数定义如下:

$$\Psi(D) : \arg \min_D \|D - \Phi\|_F^2 + \lambda \|D^T \Phi - I\|_F^2 \quad (12)$$

目标函数的第一项保证字典 D 以 Frobenius 范数逼近 r 紧框架 Φ (由定理 3 和 4 获得); 函数的第二项不同于一般的最小化原子间最大内积^[13, 16], 采用最小化原子对的内积平方和, 实现等角紧框架的构造, 从而降低字典的相干性, 同时尽量保持更新前后字典结构的整体相似特性。

本文采用梯度下降方法实现式 (12) 的求解。由矩阵的迹可知, $\|M\|_F^2 = \text{tr}(M^T M)$, 式 (12) 转化为

$$\Psi(D) = \text{tr}(\Phi D^T D \Phi^T - 2r D^T \Phi + r^2 I) + \lambda \text{tr}(\Phi^T D D^T \Phi - 2\Phi^T D + I) \quad (13)$$

$$\nabla_D \Psi(D) = 2[(D\Phi^T - rI)\Phi + \lambda(\Phi\Phi^T D - \Phi)] \quad (14)$$

$$D = D - \nabla_D \Psi(D) \cdot \text{Step} \quad (15)$$

字典更新过程具体描述见算法 1, 其中, r 为字典冗余度, Step 为迭代步长, Iteration 为迭代次数。

算法 1. 字典更新

Input: Initial dictionary D_0

Output: D

1) Initialize: Step, μ_0 , Iteration, D_0 with column-normalized;

2) For $k = 0$ to Iteration and $\mu(D) > \mu_0$;

3) Do $[USV] = \text{SVD}(D_k)$;

4) $Q_k = UV^T$;

5) $\Phi_k = \text{Normalized}(\sqrt{r}Q_k)$;

6) $\nabla_{D_k} \Psi(D_k) = 2[(D_k\Phi_k^T - rI)\Phi_k + \lambda(\Phi_k\Phi_k^T D_k - \Phi_k)]$;

7) $D_k = D_k - \nabla_{D_k} \Psi(D_k) \cdot \text{Step}$;

8) $k = k + 1$;

9) End for.

然而, 式 (12) 的优化, 即在字典更新阶段去除原子间的相干性, 其目标是逼近非相干字典, 从而忽略了式 (1) 中的稀疏表示误差, 因此, 若将算法 1 得到的非相干字典直接应用于稀疏表示会增加其表示误差, 表示性能下降。因此, 在不改变其相干性的前提下进一步优化字典以满足方程 (1), 本文采用文献 [13] 提出的方法优化字典 D , 实现信号的稀疏表示。

2.3 基于子空间旋转的字典优化算法

若 $A \in \mathbf{R}^{K \times K}$ 是任意正交阵, 则 $AA^T = I$, 对矩阵 D 实现正交变换, 字典的相干性保持不变。

$$(AD)^T AD = D^T A^T AD = G \quad (16)$$

因此, 获取非相干字典之后, 为降低信号观测值与信号估计值之间的误差, 提高信号的稀疏表示性能, 希望进一步优化字典 D 。即在满足正交约束条件下, 式 (1) 转化为

$$\arg \min_A \frac{1}{2} \|Y - ADX\|_F^2 \quad \text{s.t.} \quad A^T A = I \quad (17)$$

从子空间角度来说, 式 (17) 的求解相当于子空间旋转, 对字典 D 的优化则成为求解合适的 A , 进一步优化字典 D , 降低稀疏表示误差。

字典优化过程具体描述如算法 2 所示。

算法 2. 字典优化

Input: D, Y, X

Output: D

1) Initialize: $A = I$, nIteration, eps;

2) While $n \leq \text{nIteration}$ and $\text{norm}(Y - DX, \text{'fro'}) > \text{eps}$

3) Do $C = Y(DX)^T$;

4) $[U, S, V] = \text{SVD}(C)$;

5) $A = UV^T$;

6) $D = A * D$;

7) $n = n + 1$;

8) End while.

3 实验结果

3.1 字典的非相干性测量

本实验就本文算法 1 与 INK-SVD^[12]、IPR^[13]、IDCO^[16] 算法所构造的字典进行非相干性对比分

析. 在实验中, 字典 $D \in \mathbf{R}^{M \times K}$ 随机产生, 每个列向量归一化, 其中 $M = 64$, $K = 256$. 迭代次数 $\text{Iteration} = 30$, 最小相干性边界值 $\mu_0 = 0.1085$, $\text{Step} = 0.2$, $\lambda = 0.8$. 随迭代次数增加, 由本文算法训练字典的相干性与其他方法的对比如图 1 所示.

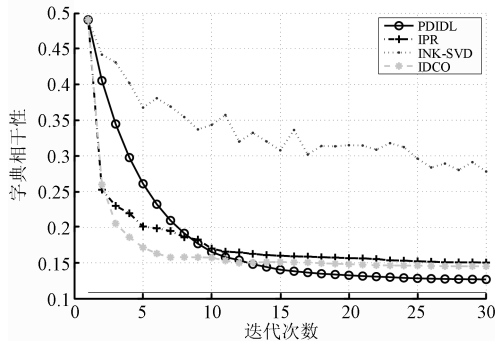


图 1 本文方法更新字典的相干性与其他方法对比

Fig. 1 Incoherence of learned dictionary with proposed method comparison to others

由图 1 可知, 虽然在迭代初期, 本文方法与 IPR 与 IDCO 方法相比, 所获取字典的相干性下降速度较慢; 但随着迭代次数的增加, 后两者过早收敛, 本文方法构造的字典的相干性仍可继续下降, 最终所获取字典的相干性均低于其他 3 种方法 (IPR、IDCO 和 INK-SVD), 更接近设定的边界值. 其主要原因在于 IPR 与 IDCO 方法直接最小化原子对内积的最大值, 虽然初始下降速度快, 但是在降低某原子相干性的同时会增加与其他原子间的相干性, 因此无法进一步收敛到设定值. 由本文方法训练的字典, 由于式 (12) 的目标函数可以逼近等角紧框架, 因而获得较好的相干性优化能力. 另外, 本实验中 INK-SVD 方法得到的字典相干性下降速度最慢, 在迭代次数相同的情况下, 其相干性最高. 其主要原因在于该方法是通过原子对两两解耦来降低字典相干性, 其降低字典相干性的效率较低.

3.2 非相干字典的稀疏表示

为了评估非相干字典的稀疏表示能力, 我们就本文方法和 K-SVD、INK-SVD、IPR 进行了对比实验. 采用欠定数学模型 $Y = DX$ 产生训练数据, 其中, 字典 $D \in \mathbf{R}^{M \times K}$ 随机产生, 服从高斯分布, 所有列向量归一化, $M = 64$, $K = 256$; X 是稀疏矩阵, $N = 1024$, 每列向量中非零元素个数为 40, 非零元素位置随机分布, 其值服从高斯分布.

本文方法实验步骤如下:

步骤 1. 采用 LARS 实现稀疏编码;

步骤 2. 基于算法 1 中步骤实现字典更新;

步骤 3. 基于算法 2 中步骤实现字典优化.

算法 1 中参数选取如下: $\mu_0 = 0.1085$, $\text{Iteration} = 30$, $\text{Step} = 0.2$, $\lambda = 0.8$; 算法 2 中参数选取如下: $n\text{Iteration} = 100$, $\text{eps} = 1\text{E}-8$; 为便于对比, K-SVD、INK-SVD、IPR 方法的迭代次数均为 30 (在 K-SVD 方法训练时, 增加了相干性的约束条件, 即相干性低于设定阈值的原子用样本中其他数据来代替). 为客观衡量本文中字典学习方法的性能, 采用学习后字典 D 的相干性和稀疏表示中 Y 的信噪比 (Signal-to-noise ratio, SNR) 值作为评价工具 (即 $\text{SNR} = 20 \lg \frac{\|Y\|_F^2}{\|Y - DX\|_F^2}$). 最终实验结果如图 2 所示, 其中图 2(a) 为利用字典进行稀疏表示时的 SNR 值, 图 2(b) 为字典相干性变化趋势, 两图横坐标均为迭代次数.

由图 2 可知, 与 K-SVD、INK-SVD、IPR 方法相比, 由本文方法获得的字典的相干性最低, 且稀疏表示时 SNR 值更高. 在迭代 10 次时, 本文方法与 IPR 方法学习字典的相干性近似相等, 而采用本文方法得到的 SNR 值为 14.67, 采用 IPR 方法得到的 SNR 值为 8.794, 这表明本文采用的最小化原子对的内积平方和比 IPR 方法中直接最小化原子最大内积更能保持字典的整体特性, 而更适合于稀疏表示. 另外, 由实验可知: K-SVD 方法无法降低字典相干性, SNR 值最低; INK-SVD 方法学习的字典相干性较高, 获取的 SNR 值也较高, 其原因在于相干性高的字典无需对原子进行解耦, 则稀疏表示性能尚可.

为更直观对上述方法所获字典的相干性和稀疏表示能力进行综合评价, 我们提出一种相对信噪比作为评价指标 $\text{SNR}/\mu(D)$. 图 3 给出了本文方法、K-SVD、INK-SVD 及 IPR 方法的相对性信噪比的性能曲线. 由图 3 易知: 采用本文的字典学习方法对于平衡字典的相干性与稀疏表示性能更有优势, 其次是 IPR、INK-SVD 方法, 而 K-SVD 方法无法实现这一目标.

3.3 语音信号的稀疏表示

为了进一步验证本文算法针对实际信号的稀疏表示性能, 利用本文方法、K-SVD、INK-SVD、IPR 方法对 SMALLBOX² 中的一段实际语音信号进行稀疏表示. 语音信号名称: music03-16 kHz, 信号频率 16 kHz. 采用矩形窗以 50% 的重叠率对原信号采样, 每个块取 256 个样本, 并将所有块重新排列作为列向量组成样本数据 $Y \in \mathbf{R}^{M \times N}$, $M = 256$, $N = 624$; 字典 $D \in \mathbf{R}^{M \times K}$, $K = 512$, 列向量归一化; X 中列向量稀疏度 $S = 12$, $\mu_{\min} = 0.0442$. 在训练数据加入服从正态分布的噪声 ($\epsilon \sim N(0, 0.2)$); 字

²<http://small.project.eu/software-data/smallbox>

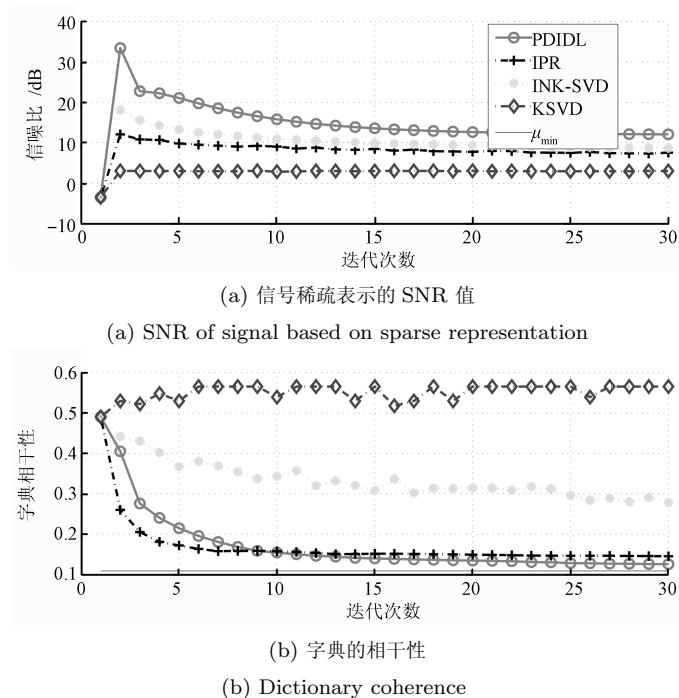


图 2 字典学习的相干性与稀疏表示性能对比

Fig. 2 Comparison of dictionary coherence and sparse representation performance the experimental results

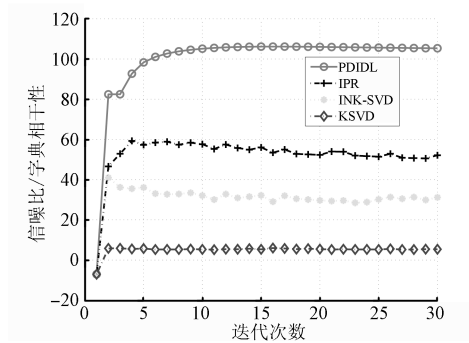


图 3 SNR 与字典相干性的比值

Fig. 3 The ratio between the SNR and dictionary coherence the experimental results

字典相干性的目标值取值范围从 0.1 到 1, 步长设置为 0.1, 所有实验分别独立进行 10 次, 计算其学习字典的相干性与稀疏表示的 SNR 值, 求相应的平均值。

初始字典分别为随机样本数据和 Gabor 字典的训练结果, 分别如图 4(a) 和图 4(b) 所示, 图 4 中给出了 4 种方法下的字典相干性与 SNR 的性能曲线。

由图 4 可知, 首先, 随着相干性的目标阈值增大, 各方法稀疏表示的 SNR 值增高, 主要原因在于字典的相干性越高, 字典学习过程中对相干性的约束则越弱, 其他学习方法就越接近于 K-SVD 方法。其次, 在字典相干性的最优设定范围内 (越小越好, 一般设为 0.1 ~ 0.35), 采用本文方法的稀疏表示 SNR

值要高于 IPR 与 INK-SVD 方法, 且信号稀疏表示性能没有出现明显衰减, 稳健性较好。K-SVD 方法的相干性 (图 4 中右上角) 无法收敛到这一区域。

因此, 本文的字典学习方法较现有 IPR、INK-SVD 和 K-SVD 方法更适合于相干性较低的字典训练, 在有效降低字典相干性的同时具有较低的稀疏表示误差。

4 结论

本文将矩阵的极分解引入到字典学习, 重新定义了优化目标函数, 提出一种新的字典学习方法。该方法不需要对格拉姆矩阵进行交替投影, 以逼近字典相干性的边界条件为前提, 直接构造等角紧框架, 对所有原子同时更新; 然后针对相干性越低的字典用于信号的稀疏表示误差越大这一问题, 通过子空间旋转方法进一步优化字典, 在保持字典相干性不发生改变的同时平衡稀疏表示性能。本文对相干性高的原子既不是简单替代, 也不是成对解耦, 提出的目标函数不仅考虑原子间的局部相似特性, 也考虑字典整体相似特性。由实验结果可知, 本文方法训练的字典较现有 K-SVD、INK-SVD、IPR 非相干字典学习方法, 可在提升稀疏表示能力的同时, 具有更低的相干性。本文研究为信号的稀疏表示与压缩感知研究提供了一种同时具有较强泛化能力与较好的稀疏表示性能的非相干字典学习方法。

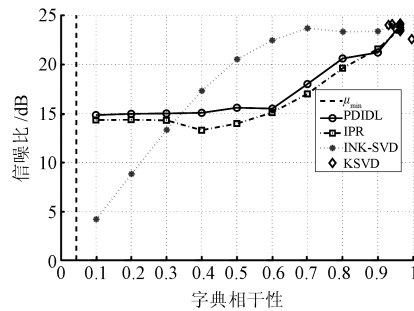
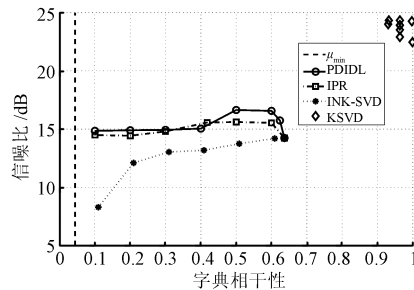
(a) 初始字典 D 为随机样本数据(a) The dictionary D is initialized random training samples(b) 初始字典 D 为 Gabor 字典(b) The dictionary D is initialized Gabor dictionary

图 4 基于不同的初始字典学习方法下的字典相干性与稀疏表示性能对比

Fig. 4 Comparison of dictionary coherence and sparse representation performance with our proposed method, K-SVD, INK-SVD, and IPR performed on the different initial dictionary for audio data

References

- Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 2006, **15**(12): 3736–3745
- Ramirez I, Sprechmann P, Sapiro G. Classification and clustering via dictionary learning with structured incoherence and shared features. In: Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA: IEEE, 2010. 3501–3508
- Candès E J, Eldar Y C, Needell D, Randall P. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 2011, **31**(1): 59–73
- Duarte-Carvajalino J M, Sapiro G. Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Transactions on Image Processing*, 2009, **18**(7): 1395–1408
- Zhang Q A, Li B X. Discriminative K-SVD for dictionary learning in face recognition. In: Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA: IEEE, 2010. 2691–2698
- Jiang Z L, Lin Z, Davis L S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D.C., USA: IEEE, 2011. 1697–1704
- Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. *SIAM Review*, 2001, **43**(1): 129–159
- Donoho D L, Huo X M. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 2001, **47**(7): 2845–2862
- Tropp J A. Greed is good: algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 2004, **50**(10): 2231–2242
- Donoho D L, Tsaig Y. Fast solution of l_1 -norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 2008, **54**(11): 4789–812
- Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 2006, **54**(11): 4311–4322
- Mailhe B, Barchiesi D, Plumbley M D. INK-SVD: learning incoherent dictionaries for sparse representation. In: Proceedings of the 2012 International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Kyoto, Japan, USA: IEEE, 2012. 3573–3576
- Barchiesi D, Plumbley M D. Learning incoherent dictionaries for sparse approximation using iterative projections and rotations. *IEEE Transactions on Signal Processing*, 2013, **61**(8): 2055–2065
- Yaghoobi M, Daudet L, Davies M E. Parametric dictionary design for sparse coding. *IEEE Transactions on Signal Processing*, 2009, **57**(12): 4800–4810
- Yaghoobi M, Daudet L, Davies M E. Structured and incoherent parametric dictionary design. In: Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Dallas, TX, USA: IEEE, 2010. 5486–5489
- Cristian R. Design of incoherent frames via convex optimization. *IEEE Signal Processing Letters*, 2013, **20**(7): 673–676
- Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2007, **53**(12): 4655–4666
- Elad M. *Sparse and Redundant Representations: from Theory to Applications in Signal and Image Processing*. New York: Springer, 2010. 95–100
- Welch L. Lower bounds on the maximum cross correlation of signals. *IEEE Transactions on Information Theory*, 1974, **20**(3): 397–399

- 20 Tropp J A, Dhillon I S, Heath R W, Strohmer T. Designing structured tight frames via an alternating projection method. *IEEE Transactions on Information Theory*, 2005, 51(1): 188–209

- 21 Horn R A, Johnson C R. *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985. 411–427



汤红忠 湘潭大学信息工程学院副教授. 湖南大学电气与信息工程学院博士研究生. 主要研究方向为压缩感知, 稀疏表示及其在图像处理与模式识别中的应用. E-mail: diandiant@126.com

(**TANG Hong-Zhong** Associate professor at the College of Information Engineering, Xiangtan University,

Ph.D. candidate at the College of Electrical and Information Engineering, Hunan University. Her research interest covers compressed sensing, sparse representation, and the application into image processing and pattern recognition.)



张小刚 湖南大学电气与信息工程学院教授. 主要研究方向为工业窑炉过程控制与模式识别.

E-mail: zhangxiaogang@126.com

(**ZHANG Xiao-Gang** Professor at the College of Electrical and Information Engineering, Hunan University.

His research interest covers process control and pattern recognition for industrial kiln.)



陈 华 湖南大学信息科学与工程学院讲师. 主要研究方向为图像处理与模式识别. 本文通信作者.

E-mail: annychen@126.com

(**CHEN Hua** Lecturer at the School of Information Science and Engineering, Hunan University. Her research interest covers image processing and pat-

tern recognition. Corresponding author of this paper.)



程 炜 湘潭大学信息工程硕士研究生. 主要研究方向为图像处理与模式识别.

E-mail: 838060390@qq.com

(**CHENG Wei** Master student at the College of Information Engineering, Xiangtan University. His research interest covers image processing and pattern recognition.)



唐美玲 湖南大学电气与信息工程学院硕士研究生. 主要研究方向为图像处理与模式识别.

E-mail: tangmeiling@hnu.edu.cn

(**TANG Mei-Ling** Master student at the College of Electrical and Information Engineering, Hunan University.

Her research interest covers image processing and pattern recognition.)