

文章编号: 1001-0920(2012)08-1237-04

联合过完备库下的信号稀疏分解

余付平¹, 冯有前², 高大化², 安芹力²

(1. 空军工程大学 导弹学院, 陕西 三原 713800; 2. 空军工程大学 理学院, 西安 710051)

摘要: 针对具有多个特征成分的复合信号在单一特征过完备库下无法实现稀疏分解的问题, 提出构建联合过完备库的思想. 联合过完备库由多个具有单一特征的过完备库联合构成, 包含复合信号中各分量信号的信息, 使得复合信号在其上具有稀疏性. 利用稀疏分解算法, 对多个复合信号在相应的联合过完备库上进行稀疏分解和信号重构, 并与单一特征过完备库的分解结果进行了对比分析. 仿真结果表明了构建联合过完备库思想的合理性和有效性.

关键词: 联合过完备库; 稀疏分解; 复合信号; 信号重构; 分解系数

中图分类号: TN957.51

文献标识码: A

Sparse decomposition over combined over-complete dictionary

YU Fu-ping¹, FENG You-qian², GAO Da-hua², AN Qin-li²

(1. Missile Institute, Air Force Engineering University, Sanyuan 713800, China; 2. School of Science, Air Force Engineering University, Xi'an 710051, China. Correspondent: YU Fu-ping, E-mail: junjingj@163.com)

Abstract: Complex signal with multiple features can not be decomposed sparsely in single feature over-complete dictionary, so the idea of building the combined over-complete dictionary is proposed. The combined over-complete dictionary is composed of many over-complete dictionaries with a single feature, and contains the information of every separate signal. The complex signal has sparsity in the combined over-complete dictionary. The decomposition result of the complex signal in a combined over-complete dictionary is contrasted with the decomposition result in an over-complete dictionary with a single feature. Simulation results show the rationality and effectiveness of the proposed idea of the combined over-complete dictionary.

Key words: combined over-complete dictionary; sparse decomposition; complex signal; signal reconstruction; decomposition coefficient

1 引言

在信号与信息处理中, 用空间变换有效地表达信号是一个重要问题. 传统的信号表示方法是基于“基”的展开, 如 Fourier 变换和小波变换等. 但这种建立在正交基上的信号分解有一定的局限性, 不是总能达到好的信号稀疏表示效果. 为了获得信号简洁稀疏的表达形式, Mallat 等人^[1]提出了基于冗余字典的稀疏分解思想.

目前, 关于信号的稀疏分解有 3 个研究方向^[2-5]: 原子的选取、过完备库(字典)的构造、求解算法设计. 其中过完备库的构造是稀疏分解中的重要环节, 关系到信号能否实现稀疏分解以及稀疏分解的效果. 过完备库的构建一般根据信号的特征构建单一特征函数的过完备库, 这种过完备库适用于信号由单一特征的

函数构成, 或者由多个特征的功能组合而成, 只有一种特征函数在该过完备库上具有稀疏性^[5-9]. 但当另外的特征函数在该过完备库上也具有大的分解系数时, 稀疏分解结果会受到影响, 特别是当组成特征函数的信号(或复合信号)成分均需要提取时, 单一特征函数的过完备库具有较大的局限性. 所以, 针对以上问题, 本文提出了构建联合过完备库的思想.

2 稀疏分解原理

给定一个集合 $\psi = \{\phi_r\}_{r \in \Gamma}$, 其元素是张成整个 N 维 Hilbert 空间 $H = \mathbf{R}^N$ 的单位矢量, 如果元素个数 $K > N$, 则集合 ψ 构成一个过完备基, 也称为字典, 其元素 ϕ_r 为由参数组 r 定义的原子, 且范数为 1. 任意信号 $S \in \mathbf{R}^N$ 在过完备原子基 $\psi \in \mathbf{R}^{N \times K}$ 上的稀疏分解问题可以描述为^[1,4,8,10]

收稿日期: 2011-03-10; 修回日期: 2011-05-18.

基金项目: 国家自然科学基金项目(61003148); 陕西省自然科学基金基础研究计划项目(SJ08F10).

作者简介: 余付平(1983-), 女, 博士生, 从事信号稀疏分解、雷达信号处理的研究; 冯有前(1960-), 男, 教授, 博士生导师, 从事信号处理、密码学等研究.

$$\min \|\alpha\|_0, \text{ 满足 } \mathbf{S} = \psi\alpha, \quad (1)$$

其中 $\|\alpha\|_0$ 为系数序列 $\{\alpha_r\}$ 中非零元素的个数. 信号在过完备库上稀疏分解的问题是: 在一定误差范围内, 如何从各种可能的解集中选择出非零元素最少的一个解向量. 然而, 该问题的求解是一个 NP 难题, 为了解决上述问题, 出现许多稀疏分解算法, 如匹配追踪 (MP) 算法、基追踪算法^[10]、框架方法^[11]和最优正交基算法^[12]等. 由于 MP 算法在逼近效果和计算复杂度方面优于其他稀疏解算法, 成为目前信号稀疏分解中较为常用的方法之一.

对于任意 N 维 Hilbert 空间的实信号 $\mathbf{S} \in \mathbf{R}^N$, 在 MP 算法下的分解过程如下^[1,2,11,13]:

Step 1: 从过完备库中选出与待分解信号最为匹配的原子 ϕ_{k_0} , 满足

$$|\langle \phi_{k_0}, \mathbf{R}_0 \rangle| = \sup_{k \in \Gamma} |\langle \phi_k, \mathbf{R}_0 \rangle|, \quad (2)$$

其中 $\mathbf{R}_0 = \mathbf{S}$. 因此信号可以分解为在最佳原子 ϕ_{k_0} 上的分量和残余两部分, 即为

$$\mathbf{R}_0 = \langle \phi_{k_0}, \mathbf{R}_0 \rangle \phi_{k_0} + \mathbf{R}_1, \quad (3)$$

其中 \mathbf{R}_1 为最佳原子对原信号进行最佳匹配后的残余.

Step 2: 对最佳匹配后的残余不断进行 Step 1 的分解过程, 即

$$\mathbf{R}_r = \langle \phi_{k_r}, \mathbf{R}_r \rangle \phi_{k_r} + \mathbf{R}_{r+1}, \quad (4)$$

其中 ϕ_{k_r} 满足

$$|\langle \phi_{k_r}, \mathbf{R}_r \rangle| = \sup_{k \in \Gamma} |\langle \phi_k, \mathbf{R}_r \rangle|. \quad (5)$$

经过 m 步分解后, 信号被分解为

$$\mathbf{S} = \sum_{r=0}^{m-1} \langle \phi_{k_r}, \mathbf{R}_r \rangle \phi_{k_r} + \mathbf{R}_m, \quad (6)$$

其中 \mathbf{R}_m 用 m 个原子的线性组合表示原始信号时所产生的误差. 在每一步分解中, 所选取的最佳原子满足式 (5), 分解的残余 \mathbf{R}_m 随着分解的进行迅速地减小. 由于 $\|\mathbf{R}_m\|$ 的衰减特性, 用少数的原子 (与信号长度相比较而言) 即可以表示信号的主要成分, 体现了信号稀疏所表示的思想.

3 联合过完备库的构建原理

假设复合信号 $\mathbf{f} = \mathbf{f}_1 + \mathbf{f}_2$, 根据信号 \mathbf{f}_1 构造具有其特征成分的过完备库 $\theta = \{\phi_i, i \in \Gamma_1\}$, $\Gamma_1 = 1, 2, \dots, L$. 根据稀疏分解理论, 信号 \mathbf{f}_1 在该过完备库上具有稀疏性. 同理, 根据信号 \mathbf{f}_2 构造包含信号 \mathbf{f}_2 特征成分的过完备库 $\varphi = \{\omega_j, j \in \Gamma_2\}$, $\Gamma_2 = 1, 2, \dots, K$, 则信号 \mathbf{f}_2 在 φ 上具有稀疏性. 信号 \mathbf{f}_1 在过完备库 φ 上不具有稀疏性, 信号 \mathbf{f}_2 在过完备库 θ 上不具有稀疏性, 即过完备 θ 和 φ 之间的相关性较小. 根据过完备库 θ 和 φ , 构建联合过完备库 $\mathbf{D} = \{\theta, \varphi\} = \{\phi_1, \phi_2, \dots, \phi_L, \omega_1, \omega_2, \dots, \omega_K\}$, 其中 θ 和 φ 称为联合过完备库 \mathbf{D} 的子原子库. 由于 θ 和 φ 之间的弱相关性和分量信号在各自子原子库上的稀疏性, 复合

信号在联合过完备库上具有稀疏性; 同时, 由于过完备库的原子模值为 1, 原子中没有包含幅值信息的参量, 从理论上讲信号稀疏分解系数的值对应各个分量信号的幅值. 参照 MP 算法, 信号 \mathbf{f} 在联合过完备库 \mathbf{D} 上进行分解的步骤如下:

Step 1: 从过完备库 \mathbf{D} 中分别选出与待分解信号 \mathbf{f} 最为匹配的两个原子, 其中一个原子 ϕ_{i_0} 属于 $\theta = \{\phi_i, i = 1, 2, \dots, L\}$, 另一个原子 ω_{j_0} 属于 $\varphi = \{\omega_j, j = 1, 2, \dots, K\}$, 且分别满足

$$|\langle \phi_{i_0}, \mathbf{R}_0 \rangle| = \sup_{i \in \Gamma_1} |\langle \phi_i, \mathbf{R}_0 \rangle|, \quad (7)$$

$$|\langle \omega_{j_0}, \mathbf{R}_1 \rangle| = \sup_{j \in \Gamma_2} |\langle \omega_j, \mathbf{R}_1 \rangle|. \quad (8)$$

其中: $\mathbf{R}_0 = \mathbf{f}$, $\mathbf{R}_1 = \mathbf{R}_0 - |\langle \phi_{i_0}, \mathbf{R}_0 \rangle| \phi_{i_0}$.

Step 2: 子原子库 θ 根据 \mathbf{f}_1 的特征信息构建, φ 体现 \mathbf{f}_2 的特征信息, 所以根据式 (7) 和 (8), 信号 \mathbf{f} 在子原子库 θ 中得到投影系数最大的原子一定包含 \mathbf{f}_1 的主要成分; 同理, 与子原子库 φ 中原子进行投影运算时, 最大投影系数对应的原子也一定包含 \mathbf{f}_2 的主要成分. 在一定收敛精度条件下, 参照 MP 算法继续迭代可以得到信号的稀疏表达式为

$$\mathbf{f} = \sum_{i=0}^{m-1} \langle \phi_i, \mathbf{R}_i \rangle \phi_i + \sum_{j=0}^{k-1} \langle \omega_j, \mathbf{R}_j \rangle \omega_j. \quad (9)$$

根据式 (9) 可以得到: 通过信号的分解系数与相对应的原子的线性组合可以重构原始信号, 恢复原始信号的特征. 由于信号 \mathbf{f}_1 和 \mathbf{f}_2 两部分特征成分相异, 且字典 \mathbf{D} 的前 L 项由包含 \mathbf{f}_1 特征成分的原子构成, 字典 \mathbf{D} 的第 $L+1$ 项到第 $L+K$ 项由包含 \mathbf{f}_2 特征成分的原子构成, 利用联合过完备库 \mathbf{D} 中前 L 项原子和与之相对应的分解系数进行线性组合能够重构分量信号 \mathbf{f}_1 , 提取 \mathbf{f}_1 的特征信息. 选取库 \mathbf{D} 中 L 项之后的原子以及与该原子相对应的分解系数的线性组合可以重构信号 \mathbf{f}_2 , 提取信号 \mathbf{f}_2 的特征信息. 本文采用均方误差 (MSE) 作为信号重构效果的评价标准, 有

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N [f(i) - f'(i)]^2. \quad (10)$$

4 仿真实验验证

选取信号 $\exp(7 \sin(t + \pi/5))$, $\exp(2 \sin(t + \pi))$, t^3 构成原始复合信号 1, 根据联合过完备库构造的理论构建过完备库 \mathbf{D} , \mathbf{D} 中的原子为

$$\left\{ \bigcup_{i=0}^{50} \bigcup_{j=0}^{20} \exp(i \sin(t + j)), \bigcup_{k=0}^{282} t^k \right\}.$$

根据待分解信号的信息, 参量 i 和 j 以间隔 1 均匀取值, 参量 k 按照 0.25 的精度进行取值. 原子个数为 2200, 采样点个数为 1000. 通过 Matlab 仿真, 分别提取 3 个分量信号的特征成分信息, 实现信号重构.

在联合过完备库 \mathbf{D} 下, 通过稀疏分解算法求得得到复合信号 1 分解系数, 如图 1 所示. 该复合信号在

联合过完备库下具有稀疏性, 且分解系数中有 3 个数远大于零, 其中两个位于过完备库的前 1000 个原子中, 另一个大于零的系数的横坐标值位于原子库后半部分原子中. 联合过完备库前 1070 个原子由包含信号 $\exp(i \sin(t + j))$ 特征的函数构成, 后 1130 个原子包含分量信号 t^3 的特征, 所以仿真结果和设计的实际信号一致. 由于分解系数中还存在近似为零的数值, 根据该系数进行信号重构存在重构误差.

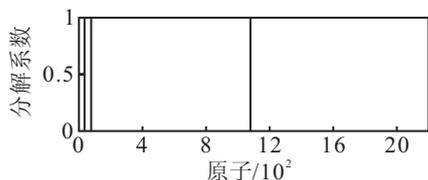


图 1 过完备库下的复合信号 1 稀疏分解系数

图 2 为原始信号 1 与重构信号 1 的对比, 原始信号及其各分量信号的重构误差分别为 $2.738 5e-14$, $3.693 5e-12$, $3.452 9e-12$, $1.653 4e-14$. 由图 2 以及信号重构误差可见, 通过联合过完备库能够实现对原始复合信号以及各分量信号无失真的重构, 恢复信号的特征成分. 通过子图之间的对比分析可以看出, 联合过完备库可以分离具有不同特征成分和具有相同特征成分的分量信号, 所以, 联合过完备库不但对具有不同特征成分的信号分解有效, 而且对具有相同特征成分的信号分解效果也较好.

为了进一步分析联合过完备库思想的合理性, 选取信号 $5 \sin(20 \times t)$, $2 \sin(50 \times t)$, t^2 组成复合信号 2, 研究该复合信号在联合过完备库下的分解效果以及信号重构效果. 针对该复合信号, 根据联合过完备库构造的理论构建过完备库 $\left\{ \bigcup_{i=0}^{50} \sin(i \times t), \bigcup_{k=0}^{200} t^k \right\}$. 参量 i 以 0.1 间隔均匀取值, 参量 k 按照 0.2 的精度进行取值. 原子个数为 2200, 采样点个数为 1000. 图 3 为在该联合过完备库下复合信号 2 的分解系数. 由图 3 可见, 复合信号在联合过完备库下的分解系数具有稀疏性, 实现了信号的稀疏分解, 且系数中非零值对应的原子位置与实际设计的信号一致; 同时, 信号稀疏分解系数中 3 个远大于零的值分别与分量信号的幅值一致, 验证了理论分析的正确性.

图 4 为复合信号 2 与各分量信号 2 的原始信号和重构信号的对比. 复合信号及各分量信号的重构误差分别为 $0.191 91$, $0.002 302$, $0.000 219 27$, $0.002 514 6$. 由此可以看出, 复合信号与分量信号重构质量均较好, 3 个分量信号的特征成分被正确识别, 分量信号实现了几乎无误差的重构. 对比图 2 和图 4 可以得到, 对于不同类型的复合信号, 通过构建合适的联合过完备库可以实现稀疏分解和信号重构.

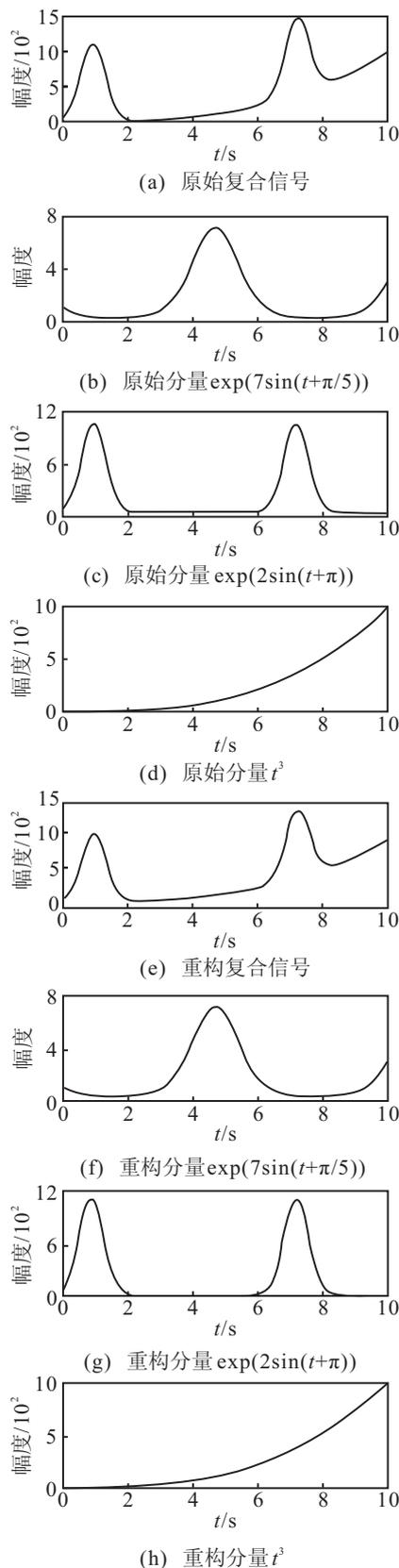


图 2 原始信号 1 和重构信号 1

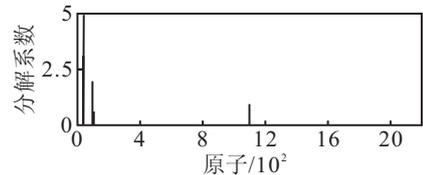


图 3 过完备库下的复合信号 2 稀疏分解系数

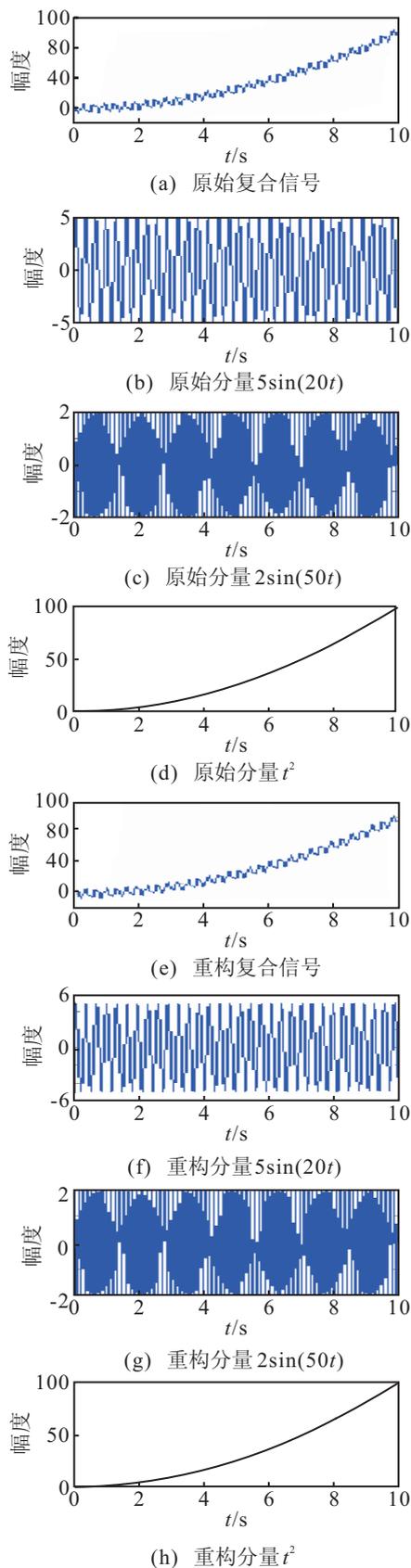


图 4 原始信号 2 和重构信号 2

为了研究构建联合过完备库的必要性和有效性,对复合信号在单一特征过完备库上进行稀疏分解,得到如图 5(a) 所示的分解系数. 由图 5 可见,复合信号在单一特征过完备库下的分解系数不具有稀疏性,信

号无法实现稀疏分解. 图 5(b) 为该信号在联合过完备库下的分解系数,可以看出分解系数明显具有稀疏性. 所以,对于包含多个特征成分的复合信号,单一特征过完备库无法兼顾多个特征成分,无法实现信号的稀疏分解,而联合过完备库包含多个特征成分,避免了特征成分的丢失,使得信号在该库下具有稀疏性.

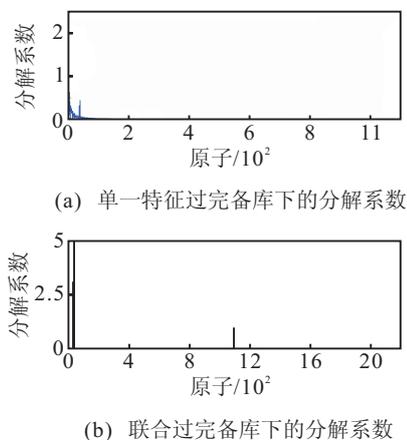


图 5 不同过完备库下的信号分解系数

5 结论

本文提出了联合过完备库的思想,与单一特征成分的过完备库不同,联合过完备库包含更丰富的待分解信号信息,克服了单一特征过完备库中特征信息不足而无法实现信号稀疏分解或稀疏分解效果不好的问题. 通过理论分析和仿真实验可以得到:具有多个特征成分的复合信号在联合过完备库上具有稀疏性,稀疏分解效果较好,而复合信号在单一特征过完备库上不具有稀疏性;对于包含多个分量信号的复合信号,联合过完备库不但能够实现稀疏分解,而且能够较好地分离和识别出各个分量信号,几乎无误差地重构复合信号及其各分量信号. 实验结果与理论分析结果一致,表明了构建联合过完备库思想的正确性和有效性. 但是,本文仅对包含两种特征函数的过完备库进行了实验分析和验证,对于含有多个特征的过完备库可以按照两种特征函数过完备库的方法进行扩展.

参考文献(References)

- [1] Mallat S, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. IEEE Trans on Signal Processing, 1993, 41(12): 3397-3415.
- [2] Michael E, Alfred M B. A generalized uncertainty principle and sparse representation in pairs of bases[J]. IEEE Trans on Information Theory, 2002, 48(9): 2558-2567.
- [3] Babaie-Zadeh M, Vigneron V, Jutten C. Sparse decomposition over non-full-rank dictionary[C]. Proc of IEEE Int Conf on Acoustics, Speech and Signal Processing. Taipei: IEEE Computer Society, 2009: 2953-2956.

(下转第 1245 页)