

从稀疏到结构化稀疏：贝叶斯方法

孙洪¹ 张智林² 余磊^{1,3}

(1. 武汉大学电子信息学院 武汉 430072; 2. Dept. Electrical and Computer Engineering, University of California, CA92093-0407, USA; 3. VisAGeS U746 Research Unit, INRIA, 35042 Rennes, France)

摘要：稀疏分解算法是稀疏表达理论和压缩感知理论中的核心问题，也是当前信号处理领域的一个热门话题。近年来，研究人员发现除了稀疏以外，如果引入稀疏系数之间的相关性先验信息，可以大大提高稀疏分解算法的精度，这种方法称为“结构化稀疏分解算法”。本文归纳和总结了从稀疏到结构化稀疏的信号模型，并且介绍了两种不同的贝叶斯稀疏(或者结构化稀疏)算法，以及从稀疏到结构化稀疏贝叶斯稀疏分解算法的扩展。同时，本文还介绍了结构化稀疏分解算法在医学信号处理和语音信号处理中的应用。

关键词：压缩感知；稀疏理论；结构化稀疏分解算法；贝叶斯压缩感知

中图分类号：TN911.7 **文献标识码：**A **文章编号：**1003-0530(2012)06-0759-15

From Sparsity to Structured Sparsity: Bayesian Perspective

SUN Hong¹ ZHANG Zhi-lin² YU Lei^{1,3}

(1. School of Electronic Information, Wuhan University, 430072 Wuhan, China;
2. Dept. Electrical and Computer Engineering, University of California, CA92093-0407, USA;
3. VisAGeS U746 Research Unit, INRIA, 35042 Rennes, France)

Abstract: Sparse decomposition algorithm is one of the hottest research topic in signal processing field and plays an important role in sparse representation and Compressive Sensing (CS). Recently, beside sparsity, the structures that describes the dependencies of sparse coefficients has been exploited to improve the accuracy of sparse decomposition algorithms. It is called structured sparse decomposition algorithms. This paper will review the sparse signal model and structured sparse signal model. After that, two sparse decomposition algorithms based on Bayesian framework are introduced and their extensions to structured sparse signals are addressed. At last, the applications of structured sparsity in medical signal processing and audio signal processing are respectively demonstrated.

Key words: Compressive Sensing; Sparsity; Structured sparse decomposition algorithms; Bayesian Compressive Sensing

1 引言

稀疏理论已经成为信号处理领域中的一个重要分支。在信号处理领域中,从信号采样^[1-3],到压缩^[4],再到建模、分类、识别^[5-10]等等,稀疏理论都取得了很大的成功。同时,稀疏理论在其他交叉领域的应用也非常广泛,如生物医学信号处理^[11-13]、无线传感器网络^[14-16]、机器学习^[17]、遥感数据处理^[18,19]等等。稀疏理论的主要思想是认为,高维空间(Ambient higher dimensional space)中的信号可以由低维信号模型(Lower dimensional signal model)来描述^[20]。信号的稀疏模型

可以描述为:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon} \quad (1)$$

其中 $\mathbf{y} \in \mathbb{R}^m$ 为信号矢量, $\mathbf{A} \in \mathbb{R}^{m \times N}$ 为稀疏映射矩阵, $\boldsymbol{\theta} \in \mathbb{R}^N$ 为稀疏矢量, $\boldsymbol{\epsilon}$ 为可能的误差项。如果定义 $\mathbf{A} = \{\mathbf{a}_i\}_{i=1, \dots, N}$, 且 $V_i = \text{span}\{\mathbf{a}_i\}$ 为基矢量 \mathbf{a}_i 展成的子空间, 那么从式(1)可以知道, 信号矢量 $\mathbf{y} \in \bigcup_{i \in K} V_i$, $K = \{i: \theta_i \neq 0\}$ 。因为 $\boldsymbol{\theta}$ 是稀疏的, 那么集合 K 中元素的个数 $|K| \ll N$, 如图 1 所示为几种典型的稀疏信号。

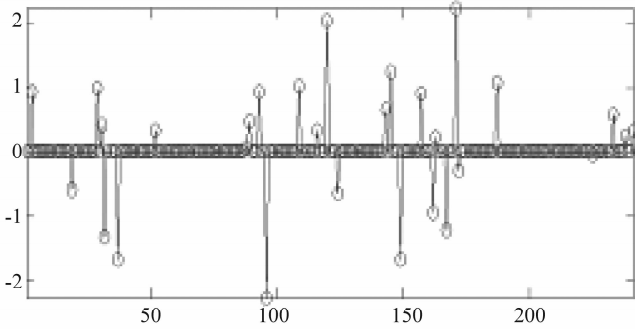
在不同的应用中, 式(1)中各个符号具体的物理意义也不尽相同。例如, 在信号的稀疏表示中, \mathbf{y} 为

原始信号, \mathbf{A} 为预先定义的稀疏表达基矩阵, 而 $\boldsymbol{\theta}$ 则是原始信号在稀疏表达基矩阵上的稀疏表达系数; 在雷达探测领域, 以波达方向估计为例, \mathbf{y} 是传感器阵列所接受的信号, \mathbf{A} 是方向矩阵 (每一列对应一个方向), $\boldsymbol{\theta}$ 的每一个非零元素所在的位置则对应着某个方向; 在模式识别领域, 以高维空间线性回归问题为例, \mathbf{y} 是因变量, \mathbf{A} 是回归系数, 而 $\boldsymbol{\theta}$ 是自变量; 在压缩感知领域, \mathbf{y} 表示压缩测量, \mathbf{A} 可以记做 $\mathbf{A} = \Phi\Psi$, 其中 $\Phi \in \mathbb{R}^{m \times M}$ 为随机感知矩阵, $\Psi \in \mathbb{R}^{M \times N}$ 为稀疏表达矩阵, 而 $\boldsymbol{\theta}$ 则是原始信号 $\mathbf{x} \in \mathbb{R}^M$ 在矩阵 Ψ 上的稀疏表达, 即 $\mathbf{x} = \Psi\boldsymbol{\theta}$.

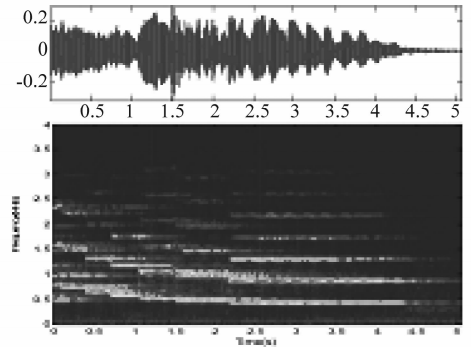
一般的, 观测维度 m 远远小于稀疏矢量的维度 N , 式(1)本身是一个病态的线性逆问题. 而信号的稀疏分解算法则是通过引入稀疏约束项来规则化对应于(1)的逆问题, 如果假设噪声项为高斯白噪声, 那么这样的稀疏规则化线性逆问题通常可以表述如下:

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \lambda\phi(\boldsymbol{\theta}) \quad (2)$$

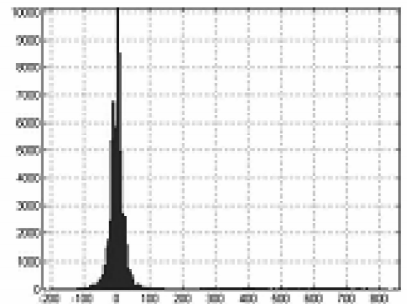
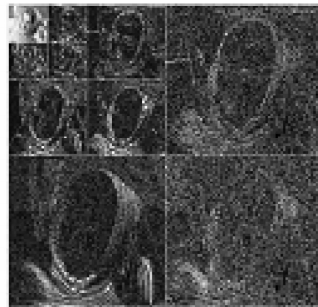
其中, $\phi(\boldsymbol{\theta})$ 为稀疏约束项, 而参数 $\lambda > 0$ 则是观测约束和稀疏约束之间的平衡参数. 通常, 稀疏约束 $\phi(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$, 其中 $\|\boldsymbol{\theta}\|_1$ 表示矢量 $\boldsymbol{\theta}$ 的 l_1 -范数. 理论证明 l_1 -范数和 l_0 -范数在适当条件下是等效的^[21].



(a) Sparse signal



(b) A musical signal and its spectrum



(c) An image, its wavelet representation and the histogram of the wavelet coefficients

图1 几种典型的稀疏信号的例子

Fig.1 Several Typical Sparse Signals

事实上, 信号的稀疏性并不是唯一的信号表示模型^[3,22-24], 例如, 在某些特定的应用中, 稀疏信号的支撑集模式是相互关联的, 即稀疏信号的非零元素具有某种隐含的结构特性^[22,25-28]. 那么这种结构信息往往可以被用来作为额外的先验信息, 用于提高稀疏表达算法的恢复精度和鲁棒性. 在压缩感知中, 这种新的扩展往往被称为基于模型的压缩感知 (Model based Compressive Sensing)^[22]. 图2给出了几种典型的结构化稀疏信号的例子: (1) 聚集结构稀疏 (Block-Structured Sparse) 信号: 这一类稀

疏信号的非零元素 (或者零元素) 呈现聚集块的形式, 在视频监控、多频带信号、基因表达、传感器网络的信源定位、MIMO 信道均衡、神经网络、核磁共振成像等很多应用中^[22,29,30], 都会经常见到; (2) 树结构稀疏 (Tree-Structured Sparse) 信号: 这一类稀疏信号的非零元素呈树状分布, 常见于小波变换系数^[28,31,32]; (3) 联合稀疏 (Joint sparsity) 信号: 这是指多个信号 (或者它们的稀疏表达系数) 的非零元素的位置保持一致, 或者指某个信号 (或者其稀疏表达系数) 的非零元素的位置在多次测量里均保持

不变。这类信号在源定位,多元回归问题中经常遇到^[33-35]; (4) 部分支撑集已知稀疏 (Partially-Known-Support Sparse) 信号: 这一类稀疏信号的某些元素的非零或零的状态 (支撑集) 已知, 例如实时动态 MRI 成像系统或者实时但像素视频成像系统中可能会出现这种情况^[36-38]。

结构化稀疏性中, 由于稀疏系数或者稀疏映射矩阵中的基矢量之间是相互依赖的, 其子空间定义也有所不同, 定义为 $U_i = \text{span}\{\mathbf{a}_j, j \in g_i\}, g_i \subset \{1, \dots, N\}$, 且 $g_i \cap g_j = \emptyset, i \neq j$ 。集合 g_i 包含了相互依赖的稀疏系数或

者基矢量的位置。那么, 信号矢量 $\mathbf{y} \in \bigcup_{i \in G} U_i, G = \{i: \theta_{g_i} \neq 0\}$, 显然集合 G 中元素的个数 $|G| < |K| \ll N$ 。因此, 结构化稀疏性减少了信号子空间的个数, 使得信号的稀疏系数更容易被精确的找到^[24,39]。不失一般性, 对应于(2)式的稀疏规则化线性逆问题, 结构稀疏规则化线性逆问题可以表述如下:

$$\boldsymbol{\theta} = \arg \min \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \lambda \phi_c(\boldsymbol{\theta}) \quad (3)$$

其中, $\phi_c(\boldsymbol{\theta})$ 为矢量 $\boldsymbol{\theta}$ 在集合 G 上的结构化稀疏约束^[40,41]。

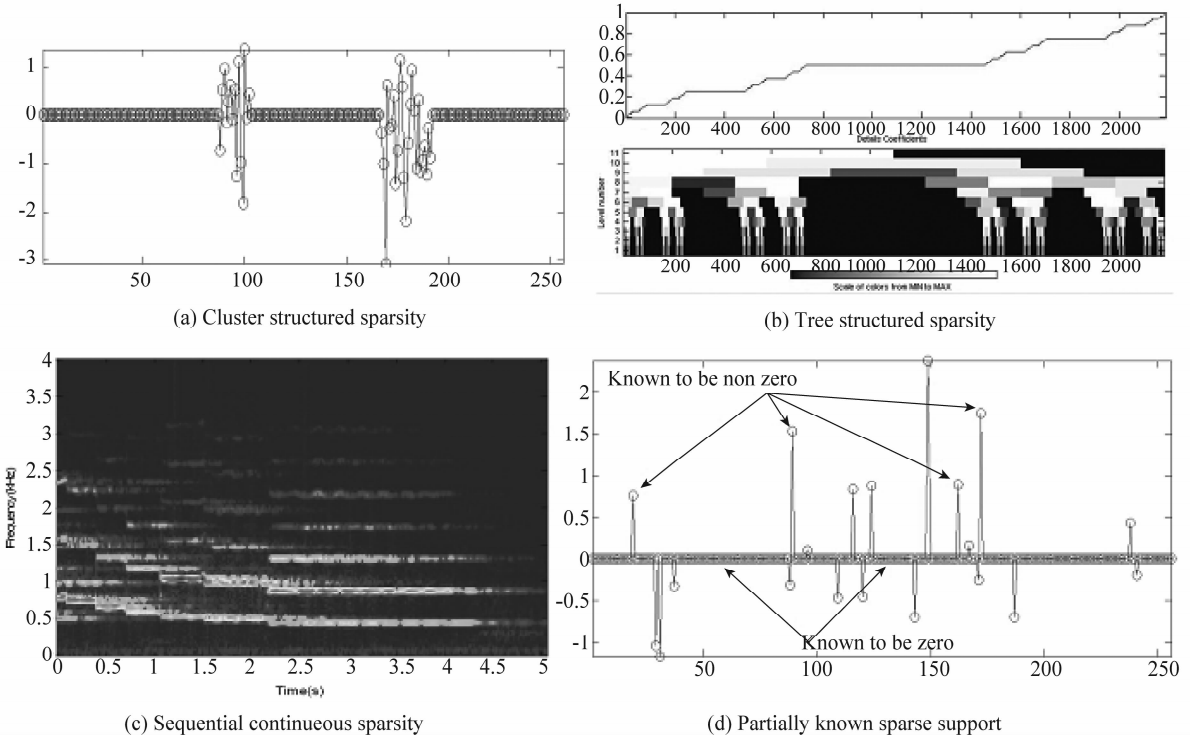


图 2 结构化稀疏信号
Fig. 2 Structured Sparse Signals

综上所述, 利用信号的结构化稀疏性能大大提高稀疏分解算法的性能。结构化稀疏信号的稀疏分解算法近年来随着压缩感知理论的发展, 这一方向的研究也引起了广泛的关注, 主要集中在如下几个方面: (1) 已知稀疏信号内在的特性, 即非零元素之间的相关性设计相应的算法^[28,42-45]; (2) 序列信号在时间轴上的连续性^[34,37,46]; (3) 部分支撑集已知的稀疏信号^[36,38]。另外, 目前在结合结构信息的稀疏表达算法中, 通常采用同匹配追踪算法类似的流程或者采用贝叶斯模型, 因为这两种方法都比较容易引入结构先验。尽管基于匹配追踪算法的结构化稀疏信号恢复算法具有严格的理论保证^[22,27], 但是这一类算法往往需要已知很多参

数, 而且这些参数往往在实际应用中很难预先得知。

贝叶斯稀疏算法的基本形式基于 Tipping 的 SBL 算法^[47]。随后 Wipf 和 Rao 等人对该类 SBL 进行了深入的理论研究。与广泛使用的基于 L1 惩罚项的算法 (比如 Lasso, BP) 相比 (以下简称 L1 算法), 此类 SBL 具有一系列显著的优势: (1) 在无噪情况下, 除非满足一些严格的条件^[48], L1 算法的全局最小点 (global minimum) 并不是真正的最稀疏的解^[47]。因此, 在一些应用中, 当真实的解是最稀疏的解, 采用 SBL 是更好的选择。(2) 当感知矩阵 (sensing matrix) 的列与列相关性很强时, L1 算法的性能会变得非常差。事实上不光是 L1 算法, 大多数已知的压缩感知算法 (比如 Approxi-

mate Message Passing 算法, Matching Pursuit 算法) 在这种情况下性能都会变得很差。相比之下, SBL 算法仍旧具有良好的性能^[49]。因此, 在雷达追踪, 脑源定位, 特征提取, 功率谱估计等一些列领域, SBL 都具备显著的优势。(3) 业已证明, SBL 算法等价于一种迭代加权 L1 最小化算法 (iterative reweighted L1 minimization), 而 L1 算法仅仅只是其第一步^[50]。Candes 等人指出, 迭代加权 L1 最小化算法更易获得真正最稀疏解^[51]。从这个角度也就不难理解 SBL 的优越性。这里我们可以将未知结构参数都看作是随机变量, 并且根据稀疏结构特性, 例如元素之间的依赖关系、又或者对于元素本身的预先判断, 对这些随机变量赋予适当的先验分布, 从而得到无参数的结构化稀疏信号的稀疏表达算法^[28, 34, 43, 46, 49, 52]。

基于贝叶斯模型, 根据结构对象的不同, 结构化稀疏性可以分别从两个方面来考虑。首先, 结构先验可能是由于稀疏系数 θ 中元素之间的相关性引起的, 这一类结构先验可以通过稀疏系数 θ 的相关函数来考虑; 其次, 结构先验还可能是由于稀疏映射矩阵 \mathbf{A} 中的基矢量 \mathbf{a}_i 的相关性引起的, 那么这一类结构先验则可以在基矢量上引入具有聚集性质的随机过程来处理。本文将从这两个方面分别介绍从稀疏到结构化稀疏的贝叶斯算法。

2 基于稀疏系数相关性的结构化稀疏贝叶斯模型

在这一章我们主要介绍一类利用稀疏系数的非零元素之间的相关性的 SBL 算法。我们已经知道, 利用信号的结构信息可以获得更好的性能^[22]。除了上一章所介绍的结构信息外, 稀疏信号非零元素的相关性也可认为是一种结构先验信息。作为一种贝叶斯算法, 这类采用参数化高斯分布为解的先验分布的 SBL 算法对利用信号的非零元素的相关性提供了更多的灵活性。最近 Zhang 和 Rao 提出了块稀疏贝叶斯学习框架 (Block Sparse Bayesian Learning, BSBL)^[46, 53]。该框架提供了一种利用解的空间结构 (spatial structure) 和时序结构 (temporal structure) 的解决方案。由其框架得到的算法在多任务学习 (multi-task learning)^[54], 生理信号的无线传输和远程监控^[53, 55], 脑源定位和脑-机接口^[56] 等许多领域获得了极大的成功。

我们首先介绍这类算法的基本形式^[47]。尽管该基本形式并没有利用解的任何结构信息, 但有助于我们去理解利用解的非零元素的相关性的 SBL 算法。为求解稀疏矢量 θ , SBL 假设 θ 中的每个元素都服从一个

参数化的均值为 0 方差为 γ_i 的高斯分布^[47]:

$$p(\theta_i; \gamma_i) = \mathcal{N}(0, \gamma_i), \quad i=1, \dots, N \quad (4)$$

其中 θ_i 表示 θ 中的第 i 个元素, γ_i 是未知的参数, 将会由算法自动估计出来。这样一种先验分布常被称为 automatic relevance 先验分布, 最初出现于神经网络领域^[57]。在算法运行中, 绝大部分的 γ_i 将会变成 0 (无噪情况下) 或者趋于 0 (有噪情况下)。SBL 通常会采用一个阈值将趋近于 0 的 γ_i 置为 0 (该阈值的大小通常和信噪比有关)。当 $\gamma_i = 0$ 时, 相应的 θ_i 则为 0。因此, γ_i 与解的稀疏程度密切相关, 也从而决定了 γ_i 的学习规则是 SBL 算法中最核心的部分。在 SBL 框架中, 噪声 ϵ 通常假设为高斯白噪声向量, 即 $p(\epsilon; \lambda) = \mathcal{N}(0, \lambda \mathbf{I})$, 其中 λ 为噪声方差。根据以上的假设, 利用贝叶斯规则很容易获得后验分布, 其也为一高斯分布。当所有的未知参数 (即 $\{\gamma_i\}_{i=1}^N, \lambda$) 都被估计出来后, θ 的最大后验估计 (Maximum A Posterior) 由这个高斯分布的均值给出。而这些未知参数可以由第二类最大似然估计 (Type II Maximum Likelihood) 获得^[17]。

除了 Tipping, Wipf 等人的 SBL 算法外, 还有其他一些 SBL 算法赋予 θ_i 的 precision 其他的分布, 或者假设 θ_i 的先验分布为一 Laplace prior^[58]。这些算法多数情况下无法证明其全局解是真正稀疏解 (即最稀疏解), 或者本身稳定性存在问题, 不能保证良好的收敛性能。值得注意的是, 赋予 θ_i 不同的先验分布并不能导致相应的 SBL 算法在实际应用中具有明显的优势。这是因为大多数实际问题都和理想的感知压缩模型相去甚远, 比如感知矩阵 (sensing matrix) 的列与列之间具有强相关性, 噪声很强, θ 并不是非常稀疏等等。在这些情况下, 不少参数的估计将会有较大的误差, 从而导致最终的解具有较大的误差。最明显的是, 绝大多数 SBL 算法对噪声方差 λ 的估计都不有效, 尤其是当感知矩阵的列与列之间具有强相关性且噪声很大的时候。而对该方差估计的准确性对 x 的估计的准确性影响非常大。Zhang 和 Rao 最近给出了噪声方差的另外一个学习规则^[46]。试验表明该学习规则可以获得更加鲁棒的效果。

事实上要想在实际中获得更好的结果, 充分利用解的自身结构信息是更加有效的策略。接下来我们将介绍如何利用解的各个元素之间的相关性来提升算法的性能。

2.1 解的空间信息和块稀疏贝叶斯学习

解的空间信息是指在稀疏向量 θ 中具有某些结构。最常见的结构是聚集结构 (cluster structure), 有的

文献也称组群结构(group structure)^[22,27,59],即

$$\boldsymbol{\theta} = [\underbrace{\theta_1, \dots, \theta_{d_1}}_{\mathbf{0}_1^T}, \dots, \underbrace{\theta_{d_{g-1}+1}, \dots, \theta_{d_g}}_{\mathbf{0}_g^T}]^T \quad (5)$$

基于这个块划分的基本压缩感知模型(即公式(1)(3))称为块稀疏模型(Block Sparse Model)。在这个模型中,解向量 $\boldsymbol{\theta}$ 可以划分为 g 个块结构(每个块结构包含的元素有多有少),而 $\boldsymbol{\theta}$ 的非零的元素则聚集在少数几个块内。基于这个模型,目前已经有了不少算法,比如 Group Lasso^[59], Block-OMP^[27], Block-CoSaMP^[22] 等等。遗憾的是,很少有算法考虑每个块内的元素之间的相关性(幅值的相关性)。为方便,以下我们称该相关性为块内相关性(Intra-Block Correlation)。

块内相关性之所以还没有引起重视,是因为在大多数情况下目前已有的算法并没有显示出其性能受到该相关性的影响。块内相关性对算法性能的影响直到最近才被 Zhang 和 Rao 通过提出块稀疏贝叶斯学习(Block Sparse Bayesian Learning, BSBL)而发现^[60],并被成功的运用到非稀疏生理信号的无线传输^[53,55]。

在 BSBL 中,每一个块 $\boldsymbol{\theta}_i$ 被假设为满足多元高斯分布:

$$p(\boldsymbol{\theta}_i) = \mathcal{N}(\boldsymbol{\theta}_i, \boldsymbol{\gamma}_i \mathbf{B}_i) \quad (6)$$

其中 \mathbf{B}_i 为一未知的正定矩阵,用于对该块内的元素之间的相关结构进行建模,而 $\boldsymbol{\gamma}_i$ 为一未知的参数,用于决定该块是否为 $\mathbf{0}$ 。类似于基本的 SBL 框架,当 $\boldsymbol{\gamma}_i = 0$, 相应的块 $\boldsymbol{\theta}_i = \mathbf{0}$ 。这样的 prior 可以认为是一种结构化的 Automatic Relevance Prior。由于 automatic relevance determination (ARD) 机制,在算法学习过程中大多数 $\boldsymbol{\gamma}_i$ 最终为 0 或者趋近于 0,从而促成了了解的块稀疏性(Block Sparsity)。同样,假设噪声服从 $p(\boldsymbol{\epsilon}; \boldsymbol{\lambda}) = \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\lambda} \mathbf{I})$ 。这样我们可以利用贝叶斯规则得到 $\boldsymbol{\theta}$ 的后验分布。利用第二类最大似然估计可以估计出各种参数,从而最终得到 $\boldsymbol{\theta}$ 的最大后验估计值。

Zhang 和 Rao 证明^[46],在无噪情况下 BSBL 的全局解即是真正的最稀疏解;而无论 \mathbf{B}_i 的值是多少都不影响这一结论。事实上, \mathbf{B}_i 的值仅仅只影响算法的局部解的性质,即算法收敛到局部解的概率。这一结论带来了极大的好处,那就是我们可以灵活采用一些策略来规范化(regularize) \mathbf{B}_i 的估计从而克服 overfitting,而无须担忧是否会影响到算法的全局解的性质。在^[46,60]中多种规范化策略被提出来。比如当每个块包含有同样数目的元素时,我们可以平均所有的 \mathbf{B}_i 得到一矩阵 \mathbf{B} ,把它作为每个 \mathbf{B}_i 的最终估计值^[60]。我们还可以进一步对 \mathbf{B} 做规范化,使 $\mathbf{B} \leftarrow \mathbf{B} + \boldsymbol{\eta} \mathbf{I}$,其中 $\boldsymbol{\eta}$ 为一正的常

数^[46]。试验表明,存在若干中规范化策略,均可获得类似的性能。

从算法层次上来说,BSBL 揭示了在标准的压缩感知试验条件下一个有意思的现象^[60,61]:当算法忽略块内相关性时(即所有的矩阵 \mathbf{B}_i 都被强制为单位矩阵),无论块内相关性是大是小,算法的性能并不发生显著的变化;当算法利用块内相关性时(即运行矩阵 \mathbf{B}_i 的学习规则),算法的性能随块内相关性的增大而提高。考虑到目前绝大多数算法都没有利用块间相关性,我们不难得到一个启示:通过改进已有的算法(比如 Group Lasso)使其可以利用块间相关性,我们可以进一步提升该算法的性能。

事实上,在^[60]中,Zhang 和 Rao 揭示了 BSBL 与 Group Lasso 等许多算法的联系。Zhang 和 Rao 证明,BSBL 的代价函数(cost function)等价于下面一迭代加权算法:

$$\boldsymbol{\theta}^{(k+1)} = \arg \min \left\| \mathbf{y} - \mathbf{A} \boldsymbol{\theta} \right\|_2^2 + \lambda \sum_{i=1}^g w_i^{(k)} \sqrt{\boldsymbol{\theta}_i^T \mathbf{B}_i^{-1} \boldsymbol{\theta}_i} \quad (7)$$

其中 k 为迭代次数, w_i 为一加权因子,其值取决于上一次迭代的结果。显然,当所有的 \mathbf{B}_i 为单位矩阵,且所有的加权 w_i 均相同且恒为一常数时,该算法即为 Group Lasso 的最常见形式。这一联系提供了如何改进已有的基于 $\sum_{i=1}^g \left\| \boldsymbol{\theta}_i \right\|_p^q$ 惩罚项的算法。感兴趣的读者可以参考^[62]了解如何改进类似 Group Lasso 的算法(即 $p = 2, q = 1$),或者参考^[63]了解如何改进迭代加权 L2 最小化算法($p = 2, q = 2$)。另外一方面,(7)也提供了如何提高 BSBL 算法的速度的一种方案。试验表明,在大多数情况下只需要迭代 3-5 次便可收敛,而每一次迭代也即运行一次 Group Lasso。考虑到目前 Group Lasso 的运算速度不断提升,采用这种迭代方式的 BSBL 算法的速度也将同步提升。

注意到在公式(6)中块的划分是已知的。在一些情况下这种划分是无法获知的。对于这种情况,文献^[60]提出了一种扩展 BSBL 框架。根据这种框架,未知块划分的情况被转化为一种简单的 BSBL 模型,而在这种 BSBL 模型中每个块的长度成为一种规范化参数(regularization parameter)。试验表明这种扩展 BSBL 框架在噪声情况下尤为有效。

值得一提的是,即使块结构未知,BSBL 框架在有些情况下也仍然适用。在^[53,55]中,BSBL 被用于恢复非稀疏的但是具有相关结构的生理信号,成功的解决了

压缩传感应用在生理信号的低能耗无线传输的瓶颈问题。在试验部分我们将给出一个代表性的例子。

2.2 解的时序信息

在一些应用中(比如源定位,雷达探测),在相继的时刻(假设为时刻 t_1, t_2, \dots, t_L)我们可以建立一系列基本的压缩传感模型: $\mathbf{y}(t_1) = \mathbf{A}\boldsymbol{\theta}(t_1) + \boldsymbol{\epsilon}(t_1)$, $\mathbf{y}(t_2) = \mathbf{A}\boldsymbol{\theta}(t_2) + \boldsymbol{\epsilon}(t_2)$, \dots , $\mathbf{y}(t_L) = \mathbf{A}\boldsymbol{\theta}(t_L) + \boldsymbol{\epsilon}(t_L)$ 。在这些模型里,感知矩阵 \mathbf{A} 均一样,每个解向量 $\boldsymbol{\theta}(t_n)$ ($n=1, \dots, L$) 的非零元素的位置也一样(但幅值可能不一样)。我们可以把这 L 个模型合并到一起,写成矩阵的形式:

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\Theta} + \mathbf{E}$$

其中 $\mathbf{Y} = [\mathbf{y}(t_1), \dots, \mathbf{y}(t_L)]$, $\boldsymbol{\Theta} = [\boldsymbol{\theta}(t_1), \dots, \boldsymbol{\theta}(t_L)]$ 。根据前面的假设, $\boldsymbol{\Theta}$ 仅有少数几行是非零行,而绝大多数行都为零行。这样一种模型称为多观测向量模型(Multiple Measurement Vector Model, MMV)^[35]。相比于仅使用一个观测向量的模型(即这 L 个基本压缩传感模型中的任意一个),MMV 模型能够更加准确的估计出非零行的位置(或者说每个解向量 $\boldsymbol{\theta}(t_n)$ 的非零元素的位置),从而更准确的获得最终的解^[27,35,61,64]。

目前已有不少针对该模型的 MMV 算法,包括基于这种 MMV 模型的 SBL 算法^[65]。遗憾的是,绝大多数算法都忽略了 $\boldsymbol{\Theta}$ 中每一个非零行内的元素之间的相关性。试验显示^[46],如果算法中不考虑这些行内的相关性,算法的性能将会极大的降低。为利用这些行内的相关性,Zhang 和 Rao 注意到块稀疏模型与 MMV 模型的联系,把 BSBL 框架用于 MMV 模型,而行内的相关结构则由 BSBL 框架中的矩阵 \mathbf{B}_i 来建模,从而得到了利用时序结构的 SBL 算法^[46]。这种利用时序结构的 SBL 算法相比于其他 MMV 算法具有明显的优势。

需要指出的是,利用时序结构信息的 MMV 算法不但可以用于信号处理领域,还可以用于机器学习领域,作为一种多任务学习(multi-task learning)算法用于特征提取。在^[54]中,利用时序结构的 SBL 算法与常见的多任务学习算法的联系被揭示出来;前者可以视为一种具有自适应核(adaptive kernels)的多任务学习算法,可以自动捕捉数据的内在结构,因而较传统的多任务学习算法以及依靠人工设定核的多任务学习算法更加优越。

2.3 解的空-时结构信息和时变信息

BSBL 框架提供了一种思路来灵活的合并解的空

间结构信息和时序结构信息。此外,利用 BSBL 框架还可以有效利用解的时变结构信息。限于篇幅我们不在这里做介绍。对这两个部分感兴趣的读者可以参考综述^[61]。

3 基于稀疏映射矩阵相关性的结构化稀疏贝叶斯模型

另外一种信号的稀疏贝叶斯模型可以采用 Bernoulli-Gaussian 模型来描述^[33,66,67]。首先,假设信号的稀疏表达 $\boldsymbol{\theta}$ 可以被分成两个部分,

$$\boldsymbol{\theta} = \mathbf{w} \circ \mathbf{z} \quad (8)$$

其中, $\mathbf{w} \in \mathbb{R}^N$ 为稀疏表达的权重参数, $\mathbf{z} \in \{0, 1\}^N$ 为稀疏表达的标号参数。那么,我们可以将式(1)改写为 $\mathbf{y} = (\mathbf{A}\mathbf{Z})\mathbf{w}$, 其中 $\mathbf{Z} = \text{diag}\{\mathbf{z}\}$ 。可以发现,虽然标号参数 \mathbf{z} 定义在稀疏矢量 $\boldsymbol{\theta}$ 上,但是其意义却代表的是稀疏映射矩阵中各个基矢量的选择与否。因此,我们可以认为这种 Bernoulli-Gaussian 模型稀疏模型是建立在基矢量 \mathbf{a}_i 上的。

假设权重参数 \mathbf{w} 满足高斯分布

$$p(\mathbf{w}_i | \alpha_i) = \mathcal{N}(\mathbf{w}_i | 0, \alpha_i^{-1}) \quad (9)$$

其中, $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1, \dots, N} \in \mathbb{R}^N$ 为超参数,且满足 Gamma 共轭先验

$$p(\alpha_i | a, b) = \Gamma(\alpha_i | a, b) \quad (10)$$

$a, b > 0$ 为模型参数。

标号参数 \mathbf{z} 可以由一个 Beta 过程来描述^[33,66],即首先假设标号参数 z_i 满足参数为 π_i 的 Bernoulli 分布:

$$\begin{aligned} p(z_i | \pi_i) &= \text{Bernoulli}(z_i | \pi_i) \\ &= \pi_i^{z_i} (1 - \pi_i)^{1 - z_i} \end{aligned} \quad (11)$$

与 Gamma-Gaussian 模型不同, Bernoulli-Gaussian 模型中,稀疏表达的非零元素的个数是由标号参数 \mathbf{z} 的非零元素的个数决定的。因此,超参数 $\pi_i \in [0, 1]$ 决定了稀疏表达中非零元素的个数。

而超参数 π_i 则满足如下 Beta 分布¹:

$$p(\pi_i | e, f) = \frac{1}{B(e, f)} \pi_i^{e-1} (1 - \pi_i)^{f-1} \quad (12)$$

其中, $B(e, f)$ 为关于参数 $e, f > 1$ 的 Beta 函数。根据 Beta 分布的性质,超参数 π_i 的期望为 $\mathbb{E}\pi_i = \frac{e}{e+f}$ 。所以, Bernoulli-Gaussian 模型中,超参数 π_i 上的 Beta 先验的两个参数 e, f 决定了稀疏表达 $\boldsymbol{\theta}$ 的稀疏度^[66]。

假设噪声项是方差为 $1/\alpha_0$ 的高斯白噪声,联合观测似然概率模型和 Bernoulli-Gaussian 稀疏模型,我们

¹ Beta 分布为 Bernoulli 分布的共轭先验。

可以用图模型来描述稀疏表达的过程,如图 3 所示。

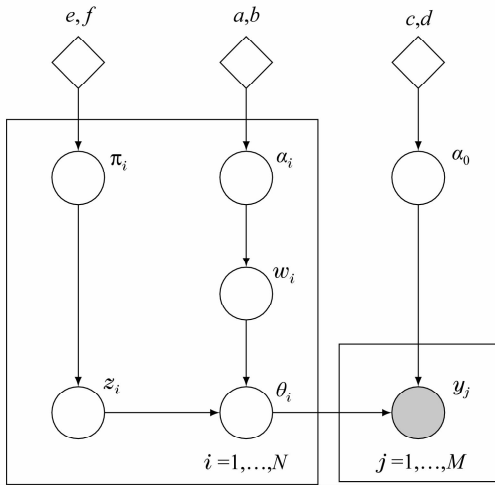


图 3 基于 Bernoulli-Gaussian 稀疏贝叶斯模型的稀疏表达
Fig. 3 Bernoulli-Gaussian Sparse Bayesian Model

现在我们已经分别定义了稀疏表达的先验概率模型,即 Bernoulli-Gaussian 模型和观测模型。那么根据贝叶斯公式,稀疏表达 θ 的后验概率可以表示为:

$$p(\theta | \mathbf{y}) = p(\mathbf{w}, \mathbf{z} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{w}, \mathbf{z}) p(\mathbf{w} | \alpha) p(\mathbf{z} | \pi) \quad (13)$$

通过计算,可以知道稀疏表达 θ 的最大后验概率估计 (MAP) 等效于求解下面的优化问题:

$$(\mathbf{w}, \mathbf{z}) = \arg \min_{\mathbf{w}, \mathbf{z}} \frac{\alpha_0}{2} \|\mathbf{y} - \mathbf{A}(\mathbf{w} \circ \mathbf{z})\|_2^2 + \log \left(\frac{1 - \mathbb{E}\pi}{\mathbb{E}\pi} \right) \|\mathbf{z}\|_0 \quad (14)$$

因此,基于 Bernoulli-Gaussian 稀疏贝叶斯模型的贝叶斯稀疏表达算法可以近似为求解 ℓ_0 优化问题^[66,67,68]。注意到,Gamma-Gaussian 稀疏模型中权重参数模型,在当 Gamma 先验参数 $a, b \rightarrow 0$ 时,趋近于 Student-t 分布,即其本身就具有稀疏促进的功能;而在 Bernoulli-Gaussian 稀疏模型的稀疏先验是通过 Beta 分布的参数引入的,这种先验不同于 CoSaMP 或者 BP 中直接给定的具体的非零元素的个数,而是通过概率的形式给出,因此更具有实际意义。同时,由于 Beta 先验负责了稀疏部分的约束,因此在式子 (10) 中,权重超参数 α_i 的 Gamma 先验参数 a, b 则不需要选择很小的值以近似稀疏分布。

3.1 基于 Bernoulli-Gaussian 模型的结构化稀疏贝叶斯模型

而基于 Bernoulli-Gaussian 模型的结构化稀疏贝叶斯模型则是通过标号参数 z 来引入结构信息。这一点与马尔可夫随机场在图像处理中的应用类似,可以将

标号参数 z 称为标号场,而权重参数 w 称为“灰度”场。那么,通过在标号场中有依赖关系的元素之间引入类似高斯马尔可夫约束,我们就可以很容易在稀疏表达中引入结构特性^[31,42,43]。但是,为了能够更容易的进行后面的贝叶斯推导,我们首先假设标号参数满足 Bernoulli 分布:

$$p(z_i | \pi_i) = \text{Bernoulli}(z_i | \pi_i)$$

那么根据稀疏表达 θ 上不同元素之间的依赖关系,我们可以将这些相互关联的元素看作是在图中聚集在一起的 Cluster。然后,考虑任意元素 θ_i 和其在 Cluster 中的邻域关系,将每三个相邻元素的关系可以分成三种不同的聚集模式:

- (1) 模式 0: 所有相邻 (依赖) 元素的标号都是 0, 那么,当前元素 θ_i 的标号在很大概率上也等于 0;
- (2) 模式 1: 邻域中只有一个 (依赖) 元素的标号是 0;
- (3) 模式 2: 所有邻域 (依赖) 元素的标号都是 1, 那么,当前元素 θ_i 的标号在很大概率上也等于 1。

对于不同的聚集模式,我们通过一个模式选择过程选择不同的参数的 Bernoulli 分布:

$$\pi_i = \begin{cases} \pi_i^{(0)}, & \text{模式 0;} \\ \pi_i^{(1)}, & \text{模式 1;} \\ \pi_i^{(2)}, & \text{模式 2。} \end{cases} \quad (15)$$

其中,三个不同的超参数 $\pi_i^{(0)}, \pi_i^{(1)}, \pi_i^{(2)}$ 分别服从 Beta 分布 (Beta 分布为 Bernoulli 分布的共轭先验):

$$\begin{aligned} p(\pi_i^{(0)} | e^{(0)}, f^{(0)}) &= \text{Beta}(\pi_i^{(0)} | e^{(0)}, f^{(0)}); \\ p(\pi_i^{(1)} | e^{(1)}, f^{(1)}) &= \text{Beta}(\pi_i^{(1)} | e^{(1)}, f^{(1)}); \\ p(\pi_i^{(2)} | e^{(2)}, f^{(2)}) &= \text{Beta}(\pi_i^{(2)} | e^{(2)}, f^{(2)}) \end{aligned} \quad (16)$$

那么,我们可以通过选择不同的参数 $e^{(k)}, f^{(k)}$ 来引入与对应模式适应的超参数先验,例如对于模式 0,可以选择 $e^{(0)} < f^{(0)}$ 使得很大概率上得到比较小的 $\pi^{(0)}$,从而驱使当前元素的标号 z_i 在很大概率上为 0。

综上所述,我们可以将基于 Bernoulli-Gaussian 模型的结构化稀疏模型用一个图模型来描述,如图 4 所示。通过隐变量 z 引入稀疏系数结构先验的方法,文献^[28,31]提出了树结构稀疏信号的压缩感知恢复算法;文献^[42,69]则提出了聚集结构稀疏信号的压缩感知恢复算法。由于采用了分层贝叶斯模型,并且未知参数通过引入合理的先验分布,使得最后得到的稀疏分解算法是完全无参数的。

在基于 Bernoulli-Gaussian 模型的结构化稀疏模型中,可以发现, Bernoulli-Beta 模型则被用来引入结构信息,这一点与 Bernoulli-Gaussian 稀疏模型不同。因此,

在 Bernoulli-Gaussian 模型的结构稀疏模型中,其稀疏部分的约束仍然是由 Gamma-Gaussian 模型部分,即式子(9)和(10),那么参数 a, b 应选取较小的值。

假设噪声项是方差为 $1/\alpha_0$ 的高斯白噪声。根据贝叶斯公式,稀疏表达 θ 的最大后验概率估计 (MAP) 近似等效于优化下面的问题:

$$(\mathbf{w}, \mathbf{z}) = \arg \min_{\mathbf{w}, \mathbf{z}} \frac{\alpha_0}{2} \|\mathbf{y} - \mathbf{A}(\mathbf{w} \circ \mathbf{z})\|_2^2 + \psi(a, b) \sum_{i=1}^N \ln \left(b + \frac{w_i^2}{2} \right) + \phi(\mathbf{z}) \quad (17)$$

与 Bernoulli-Gaussian 稀疏模型的最大后验估计(14)不同,式子(17)中第一项是观测匹配项;第二项为 Gamma-Gaussian 模型引入的稀疏约束项,因此参数 a, b 需要选择较小的值,使其近似与稀疏分布,另外 $\psi(a, b) = a + \frac{1}{2}$;第三项 $\phi(\mathbf{z})$ 为结构约束项,其中

$$\phi(\mathbf{z}) = \sum_{i=1}^N \phi(z_i | z_{\mathcal{J}_i})$$

且 $z_{\mathcal{J}_i}$ 表示标号参数 z_i 的依赖项, $\phi(z_i | z_{\mathcal{J}_i})$ 可以写成与(15)-(16)式对应的选择函数。

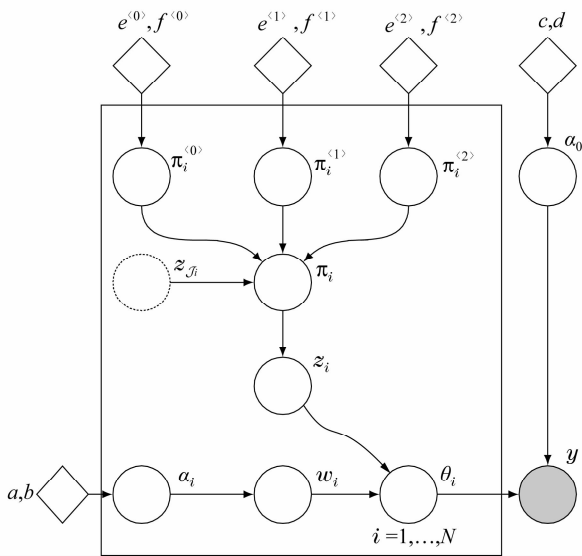


图4 基于 Bernoulli-Gaussian 模型的结构化稀疏模型

Fig.4 Bernoulli-Gaussian Structured Sparse Bayesian Model

但是,当基于上述“模式选择”的结构稀疏分解算法在当聚集块中元素过多时,就需要赋予很多种不同的模式,同时对于参数 $e^{(k)}, f^{(k)}$ 的选择也会随之变得比较复杂。因此,在文献^[69]中,我们还提出了一种更简单的结构稀疏模型,这种模型基于局部 Beta 过程,将位于同一个聚集组中的元素的标号看作是由一个局部 Beta 过程产生的,由于 Beta 过程本身具有聚类性

质^[33,70],因此不需要再使用“模式选择”过程来促进聚集结构。具体算法,可以参见文献^[69],这里就不详述了。

3.2 基于 Boltzmann Machine 的结构稀疏模型

文献^[71]提出了一种更一般的模型,通过 Boltzmann Machine (BM) 从标号场 \mathbf{z} 引入结构稀疏,这里假设标号场 \mathbf{z} 满足 BM 分布,即

$$p(\mathbf{z}) = \frac{1}{Z} \exp \left(\mathbf{t}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \mathbf{D} \mathbf{z} \right) \quad (18)$$

其中, \mathbf{t} 为 BM 参数(可以通过变换与 Bernoulli 分布的参数 π 等效), Z 为归一化常数, \mathbf{D} 为相关矩阵。可以发现,通过定义不同的相关矩阵 \mathbf{D} ,可以引入不同的稀疏结构。通常,矩阵 \mathbf{D} 可以根据数据的性质预先定义^[71],或者通过学习的方式从数据中估计^[72]。详细算法可以参见文献^[71]。

4 应用

4.1 非稀疏的生理信号的压缩传感及其在医学远程监控中的应用

经由无线体域网 (Wireless Body-Area Network, WBAN) 的生理信号的远程监控是目前医疗通讯领域的一个主要发展方向。最近开始有学者提出把压缩传感技术用于这一领域^[73-75]。压缩传感技术用于这一领域有很明显的优势:(1)当采用元素仅为 0 或者 1 的稀疏矩阵 (binary sparse matrix) 为传感矩阵时,压缩传感可以比传统的小波压缩技术更加减少无线体域网的能量损耗^[75]。而减少能量损耗是无线体域网研究的一个核心问题^[76,77]。(2)从压缩质量上来看,压缩传感和小波压缩有类似的压缩率和恢复质量。

但是压缩传感技术在这一领域取得的成果仅仅局限在及少数几种特别稀疏的生理信号(比如高信噪比的成人心电图信号),或者局限在少数几种特殊用途上。压缩传感技术还不能用于更多的生理信号(比如胎儿心电图信号,脑电图信号),以及还不能用于更为广泛的用途。这主要是因为:(1)绝大多数的生理信号是非稀疏的信号,即使被投影到变换域上(比如小波域)仍旧不足够稀疏^[55]。尽管这些生理信号可以认为是一种可压缩 (compressive) 的信号(即:虽然信号的所有元素都非零,但是只有少数一些元素显著非零,而绝大部分元素接近为零),但是压缩传感算法只能保证恢复信号中的显著的非零元素,而接近为零的元素无法准确恢复。但是对于医学远程监控来说,生理信号中一些接近为零的元素常常是重要的诊断指标。比如在

胎儿心电图,感兴趣的胎儿心跳信号常常非常微弱,接近于零,甚至很难被肉眼察觉。(2)生理信号被传输到远程终端之后,通常还需要经过一系列的信号处理和模式识别。比如当多通道脑电信号被传输到远程终端之后,通常还需要进行独立分量分析^[78],从脑电信号中提取感兴趣的独立分量,然后再进行模式识别和医疗诊断。这些应用要求压缩传感算法能够完整的把生理信号压缩并恢复出来,而不损失任何细微的波形特征,否则无法进行后继的信号处理和医疗诊断^[53]。

以上两个因素要求压缩传感算法能够以极高质量恢复非稀疏信号或者可压缩信号。而我们就所知,除了 BSBL 算法外,目前还没有任何一个压缩传感算法做到这一点。接下来我们将给出 BSBL 算法解决这一瓶颈问题的几个代表性例子。这些例子取自^[53,55]。

首先给出用 BSBL 算法实现胎儿心电信号的压缩传感。作为示例,我们选用 BSBL 框架中的 BSBL-BO 算法^[60]。

图 5(a) 给出一段采集到的原始胎儿心电图信号(数据来自:<http://www.esat.kuleuven.ac.be/sista/dai-sy>)。其中两个最明显的峰为母亲心电信号的 QRS 成分。另外有两个小峰分别出现在第 80 个采样点和第 200 个采样点附近,他们分别对应胎儿心电信号的 QRS 成分,也是我们真正感兴趣的成分。令这段原始信号为 θ 。我们任意生成一个 125×250 维的 binary 稀疏传感矩阵,记为 A 。 A 中每一列仅有 15 个元素为 1,其余元素都为 0。由 θ 和 A 便得到了压缩的信号 $y = A\theta$ 。 y 经由无线域网传到智能手机并经过互联网最终传输到远程医疗终端。在该终端,BSBL-BO 算法由压缩的信号 y 和共享的传感矩阵 A 恢复原始的胎儿心电信号 θ 。图 5(b) 是恢复的结果。显然,胎儿心电信号的 QRS 成分被很好的恢复出来。图 5(c) 给出当 BSBL-BO 不利用块内相关性的结果。此时,胎儿心电信号的 QRS 成分不能完整的恢复出来,这说明了利用块内相关性的重要性。

在这个例子中,BSBL-BO 算法假设 θ 可以划分为 10 个块,每个块都包含 25 个元素。显然这种块划分并不符合 θ 的真实块结构(见图 5(a))。但是事实上 BSBL 算法并不需要这种块划分和信号的真实块划分一致;这种假设的块划分可以视为一种规范化(regularization)。尽管假设的不同块划分会产生性能的一些微小变化,但是算法的性能对块划分并不敏感^[53,60]。这一现象其实也可以借助贝叶斯估计理论从理论上解释。

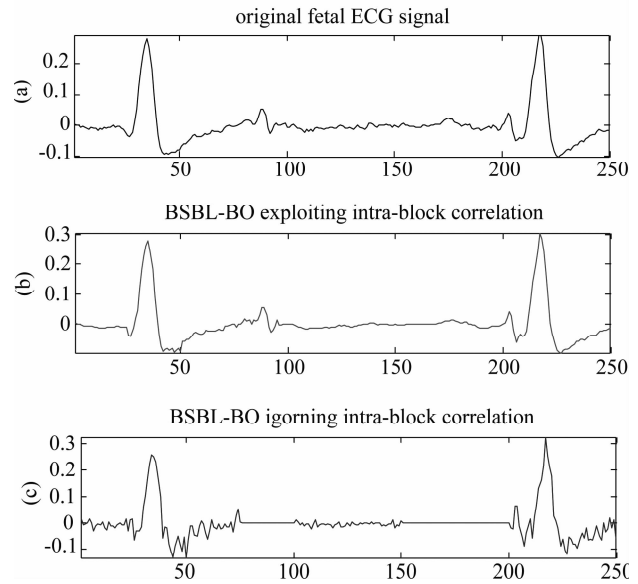


图 5 原始的胎儿心电图(a)和用 BSBL-BO 算法恢复的胎儿心电图。其中(b)是 BSBL-BO 利用块内相关性的结果,(c)是不利用块内相关性的结果
Fig. 5 The original Fetal ECG signal (a) and its reconstruction from compressed measurements by BSBL-BO considering intra-block correlation (b), and ignoring intra-block correlation (c)

作为比较,我们挑选了六种代表性的压缩传感算法来恢复该原始信号。它们分别是 CoSaMP^[79], Elastic Net^[80], SLO^[81], Block-OMP^[27], StructOMP^[82], 和 BM-MAP-OMP^[71]。其中前三种不利用任何结构信息,后三种利用了块结构信息。具体的参数设置见^[53]。图 6 给出了这些算法恢复的结果。显然,所有的算法都不能成功恢复出胎儿心跳信号的 QRS 峰。

这个例子清晰的说明了 BSBL 算法相比于其他算法的显著优势,同时也说明了利用块内相关性的好处。其实 BSBL 算法的优势远不止这些。我们前面已说过,生理信号在传输到远程终端后,通常还要经过一系列的信号处理和模式识别。一旦恢复的信号中所引入的畸变稍大,则会导致后继的信号处理产生极大的误差。为了研究 BSBL 算法是否真正有助于远程监控和医疗诊断,在^[53,55]中分别对胎儿心电信号和脑电信号进行了一系列试验。试验中,BSBL-BO 首先恢复出多通道的生理信号(胎儿心电图,脑电信号等),然后在恢复的多通道信号进行高级信号处理(比如独立分量分析),得到最终信号。然后,对原始采集到的多通道生理信号进行同样的高级信号处理,得到另外一组最终信号。最后比较两次获得的最终信号。结果显示这两次获得的最终信号几乎一模一样。这个结果说明

了BSBL-BO算法在恢复非稀疏的生理信号中,可以保证极高的恢复质量,对远程监控和医疗诊断具有实际

的价值。我们预期在不远的将来诸如BSBL等各种结构化稀疏算法在这一领域获得广泛应用。

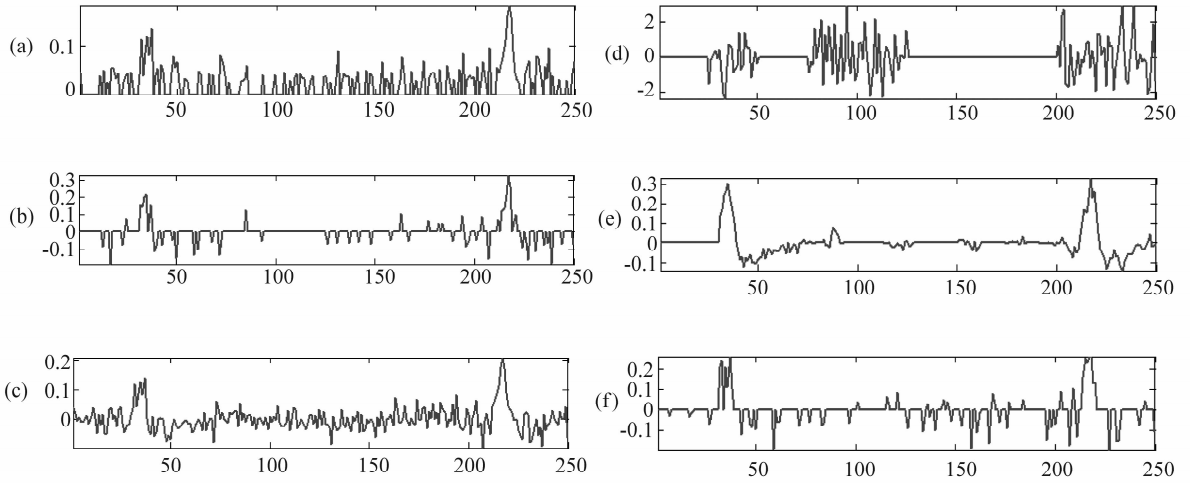


图6 六种代表性的算法恢复的结果:(a) CoSaMP, (b) Elastic Net, (c) SLO, (d) Block-OMP, (e) StructOMP, (f) BM-MAP-OMP

Fig. 6 Reconstruction results via the following six typical algorithms:

(a) CoSaMP, (b) Elastic Net, (c) SLO, (d) Block-OMP, (e) StructOMP, (f) BM-MAP-OMP

4.2 语音信号恢复

人或者音乐器件发出的声音信号可以看作是多个频率的正弦信号的混叠构成的,因此虽然语音信号在时域中并不稀疏,但是通过离散余弦变换(MDCT),我们可以将语音信号变换到频率域,使得在频率域该信号是稀疏的,传统的MP3压缩技术就是基于这个原理。同时,由于频率分辨率或者语音信号本身的特性,导致语音信号在变换域(频率域)具有时频聚类特性,如图7所示。文献^[25]基于这种性质提出了分块阈值语音去噪算法,文献^[42]则利用聚集稀疏性提高了语音压缩感知恢复算法的恢复性能。因此,我们可以将语音信号认为是一类聚集结构稀疏信号,本节将在语音压缩感知中考虑聚集结构稀疏性质,采用基于Bernoulli-Gaussian聚集结构稀疏恢复算法对语音信号的恢复性能,从而说明结构化稀疏性对语音压缩感知恢复性能的提升。

压缩感知可以用下面的线性方程来描述^[2]:

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\epsilon} \quad (19)$$

其中, $\Phi \in \mathbb{R}^{m \times M}$ 为感知矩阵,通常可以定义为高斯随机矩阵; $\mathbf{y} \in \mathbb{R}^m$ 为压缩测量, $\boldsymbol{\epsilon}$ 为噪声干扰项, $\mathbf{x} \in \mathbb{R}^M$ 为原始信号,且假设 \mathbf{x} 是稀疏信号,可以由矩阵 Ψ 稀疏表达,即 $\mathbf{x} = \Psi \boldsymbol{\theta}$ 。联合(19)式,可以得到

$$\mathbf{y} = \Phi \Psi \boldsymbol{\theta} + \boldsymbol{\epsilon} = \mathbf{A} \boldsymbol{\theta} + \boldsymbol{\epsilon} \quad (20)$$

那么,当感知矩阵 Φ 与稀疏表达矩阵 Ψ 的不相干性很

大时^[83],那么可以证明矩阵 $\mathbf{A} = \Phi \Psi$ 满足RIP性质,且稀疏表达 $\boldsymbol{\theta}$ 可以被唯一的确定。

在语音压缩感知中,其稀疏表达矩阵 Ψ 定义为DCT变换矩阵,那么根据语音信号在频域的聚集性质,可以提高语音信号的恢复性能。如图8所示,CluSS-MCMC^[42], CluSS-VB^[84], MBSC-LBP^[69]算法在稀疏的基础上引入了聚集结构性质,而其他四种算法则仅仅考虑了语音在频域的稀疏性,试验结果表明,同时考虑信号的稀疏性和结构可以很大的提高信号的恢复精度。

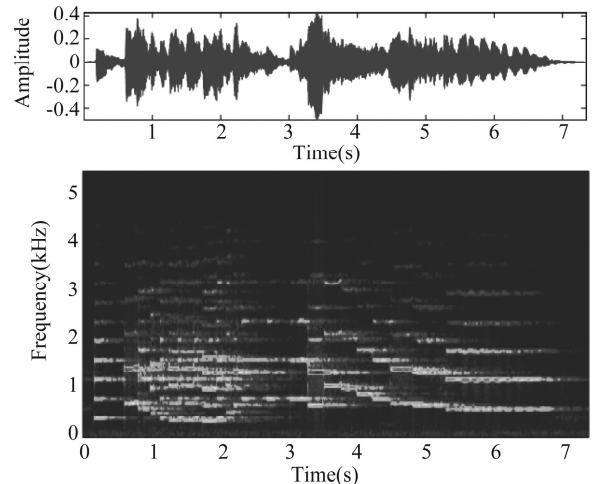


图7 一段音乐信号及其语谱图

Fig. 7 A clip of musical signal and its spectrum

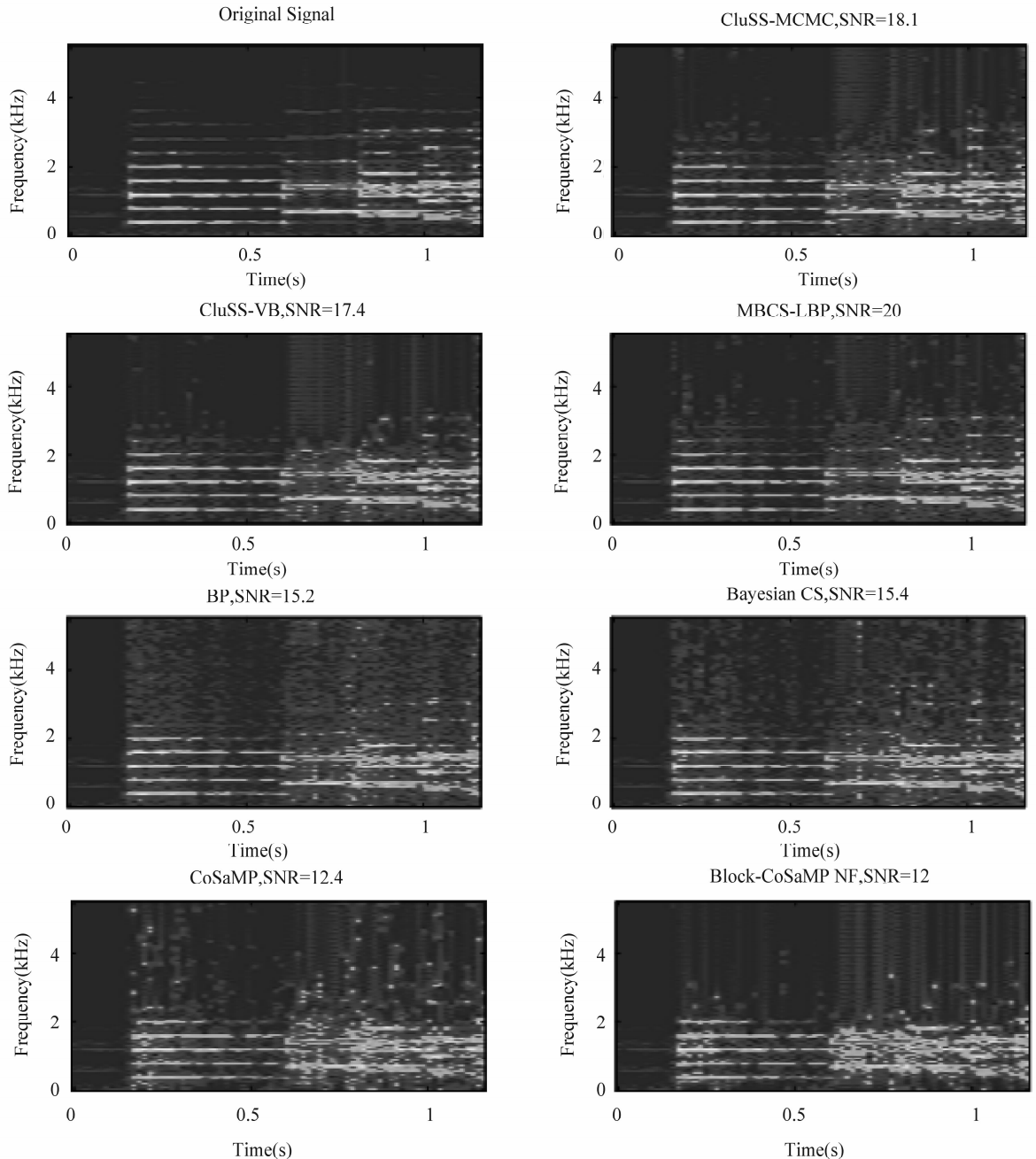


图 8 不同算法从压缩测量中恢复原始语音信号的语谱图: CluSS-MCMC^[42], CluSS-VB^[84], MBCS-LBP^[69], BP^[85], Bayesian CS^[17], CoSaMP^[79], Block-CoSaMP^[22]

Fig. 8 The spectrums of reconstruction results from the following algorithms: CluSS-MCMC^[42], CluSS-VB^[84], MBCS-LBP^[69], BP^[85], Bayesian CS^[17], CoSaMP^[79], Block-CoSaMP^[22]

5 总结

近年来,结构化稀疏模型逐渐引起了广泛的关注。本文从信号的稀疏模型出发,论述了在稀疏模型基础上的结构化稀疏模型。结构化稀疏模型不仅能描述信

号的稀疏性,而且能准确的描述稀疏系数(或者稀疏映射矢量)之间的相关性。因此,基于结构化稀疏模型,在信号压缩方面,我们能进一步提高信号的压缩率;在波达方向估计(基于稀疏理论的盲源分离技术)上,能进一步提高估计的准确性;在压缩信号的恢复方面,能

提高恢复算法的精度以及鲁棒性。而结构化稀疏分解算法则是核心内容,本文从贝叶斯方法的角度,分别从稀疏系数之间和稀疏映射矢量之间的相关性出发,介绍了基于两种不同的稀疏贝叶斯模型的结构化稀疏贝叶斯模型。并且,我们将这两种算法分别应用于医学信号处理和语音信号处理中,取得了很好的性能。

参考文献

- [1] Donoho D L. Compressed Sensing[J]. IEEE T. Inform. Theory. ,2006,52(4):1289-1306.
- [2] Candès E J, Wakin M B. An Introduction To Compressive Sampling[J]. IEEE Signal. Proc. Mag. ,2008,25(2):21-30.
- [3] Unser M. Sampling-50 Years after Shannon[J]. Proceedings of the IEEE,2000,88(4):569-587.
- [4] DeVore R A, Jawerth B, Lucier B J. Image compression through wavelet transform coding[J]. IEEE Transactions on Information Theory,1992,38(2):719-746.
- [5] Fadili M J, Starck J L, Bobin J, et al. Image decomposition and separation using sparse representations: An overview[J]. Proceedings of the IEEE, IEEE,2010,98(6):983-994.
- [6] Bertalmio M, Vese L, Sapiro G, et al. Simultaneous Structure and Texture Image Inpainting[J]. IEEE T. Image. Process. ,2003,12(8):882-889.
- [7] Elad M, Starck J-L, Querre P, et al. Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA)[J]. Applied and Computational Harmonic Analysis,2005,19(3):340-358.
- [8] Huang K, Aviyent S. Sparse Representation for Signal Classification[C]//NIPS. 2007:609-616.
- [9] Huang T. Linear spatial pyramid matching using sparse coding for image classification[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009:1794-1801.
- [10] Wright J, Yang A Y, Ganesh A, et al. Robust Face Recognition via Sparse Representation[J]. IEEE T. Pattern. Anal. ,2009,31(2):210-227.
- [11] Lustig M, Donoho D L, Pauly J M. Sparse MRI: The application of compressed sensing for rapid MR imaging [J]. Magnetic Resonance in Medicine, Wiley Online Library,2007,58(6):1182-1195.
- [12] Escoda O D, Granai L, Lemay M, et al. Ventricular and Atrial Activity Estimation Through Sparse Ecg Signal Decompositions[C]//2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings. IEEE,2:II-1060-II-1063.
- [13] Lustig M, Donoho D L, Santos J M, et al. Compressed Sensing MRI [J]. IEEE Signal Processing Magazine, 2008,25(2):72-82.
- [14] Quer G, Masiero R, Munaretto D, et al. On the interplay between routing and signal representation for Compressive Sensing in wireless sensor networks[C]//2009 Information Theory and Applications Workshop. IEEE,2009:206-215.
- [15] Haupt J, Bajwa W U, Raz G, et al. Toeplitz compressed sensing matrices with applications to sparse channel estimation[J]. IEEE T. Inform. Theory. , IEEE,2010,56(11):5862-5875.
- [16] Bajwa W, Haupt J, Sayeed A, et al. Compressive wireless sensing[C]//Proceedings of the fifth international conference on Information processing in sensor networks-IPSN '06. New York, New York, USA:ACM Press,2006:134.
- [17] Tipping M E. Sparse Bayesian Learning and the Relevance Vector Machine[J]. J. Mach. Learn. Res. ,JMLR. org, 2001,1:211-244.
- [18] Ma J. Single-Pixel Remote Sensing[J]. IEEE Geosci. Remote. S. ,2009,6(2):199-203.
- [19] Dai D, Yang W. Satellite Image Classification via Two-Layer Sparse Coding With Biased Image Representation[J]. IEEE Geoscience and Remote Sensing Letters,2011,8(1):173-176.
- [20] Wakin M B. The Geometry of Low-Dimensional Signal Models[D]. Stanford University,2006:167.
- [21] Candès E J. The Restricted Isometry Property and Its Implications for Compressed Sensing[J]. Comptes Rendus Mathématique, Elsevier,2008,346(9-10):589-592.
- [22] Baraniuk R G, Cevher V, Duarte M F, et al. Model-Based Compressive Sensing[J]. IEEE T. Inform. Theory. ,2010,56(4):1982-2001.
- [23] Vetterli M, Marziliano P, Blu T. Sampling Signals with Finite Rate of Innovation[J]. IEEE T. Signal. Proces. , IEEE,2002,50(6):1417-1428.
- [24] Blumensath T. Sampling and reconstructing signals from a union of linear subspaces[J]. Arxiv preprint arXiv:0911.3514,2009.
- [25] Yu G, Mallat S, Baccry E. Audio Denoising by Time-Frequency Block Thresholding[J]. IEEE T. Signal. Pro-

- ces. ,2008,56(5):1830-1839.
- [26] Fevotte C, Godsill S J J. A Bayesian Approach for Blind Separation of Sparse Sources[J]. IEEE Transactions on Audio, Speech and Language Processing, IEEE, 2006, 14(6):2174-2188.
- [27] Eldar Y C, Kuppinger P, Bolcskei H. Block-Sparse Signals: Uncertainty Relations and Efficient Recovery[J]. IEEE T. Signal. Proces. ,2010,58(6):3042-3054.
- [28] He L, Carin L. Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing[J]. IEEE T. Signal. Proces. ,2009,57(9):3488-3497.
- [29] Cevher V, Hegde C, Duarte M F, et al. Sparse Signal Recovery Using Markov Random Fields[C]//NIPS. 2008.
- [30] Cevher V, Indyk P, Hegde C, et al. Recovery of Clustered Sparse Signals from Compressive Measurements [C] // Int. Conf. on Sampling Theory and Applications (SAMP-TA). 2009.
- [31] He L, Chen H, Carin L. Tree-Structured Compressive Sensing With Variational Bayesian Analysis [J]. IEEE Signal. Proc. Let. ,2010,17(3):233-236.
- [32] Baraniuk R G. Optimal tree approximation with wavelets [C]//SPIE proceedings series. 1999:196-207.
- [33] Paisley J, Carin L. Nonparametric Factor Analysis with Beta Process Priors [C] //ICML. New York, NY, USA: ACM, 2009:777-784.
- [34] Ji S, Dunson D, Carin L. Multitask Compressive Sensing [J]. IEEE T. Signal. Proces. ,2009,57(1):92-106.
- [35] Cotter S F, Rao B D, Engan K, et al. Sparse solutions to linear inverse problems with multiple measurement vectors [J]. IEEE T. Signal. Proces. , IEEE, 2005, 53(7):2477-2488.
- [36] Vaswani N, Lu W. Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support [J]. IEEE T. Signal. Proces. ,IEEE,2010,58(9):4595-4607.
- [37] Vaswani N. LS-CS-Residual (LS-CS): Compressive Sensing on Least Squares Residual[J]. IEEE T. Signal. Proces. ,2010,58(8):4108-4120.
- [38] Jacques L. A Short Note on Compressed Sensing with Partially Known Signal Support[J]. Signal Processing, 2010, 90(12):3308-3312.
- [39] Lu Y M, Do M N. A Theory for Sampling Signals from a Union of Subspaces[J]. IEEE T. Signal. Proces. ,IEEE, 2008,56(6):2334-2345.
- [40] Jenatton R, Audibert J Y, Bach F R. Structured variable selection with sparsity-inducing norms[J]. Arxiv preprint arXiv:0904.3523, 2009.
- [41] Bach F R. Structured sparsity-inducing norms through submodular functions[J]. Arxiv preprint arXiv:1008.4220, 2010.
- [42] Yu L, Barbot J-P P, Zheng G, et al. Bayesian compressive sensing for cluster structured sparse signals[J]. Signal Processing, 2010, 92(1):259-269.
- [43] Yu L, Sun H, Barbot J-P, et al. Bayesian Compressive Sensing for Clustered Sparse Signals [C] // Proc. ICASSP. 2011:3948-3951.
- [44] Duarte M F, Wakin M B, Baraniuk R G. Wavelet-domain Compressive Signal Reconstruction using a Hidden Markov Tree Model[C] //Proc. ICASSP. 2008:5137-5140.
- [45] Eldar Y C, Bolcskei H. Block-sparsity: Coherence and Efficient Recovery[C] //Proc. ICASSP. 2009:2885-2888.
- [46] Zhang Z, Rao B. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning [J]. IEEE Journal of Selected Topics in Signal Processing, IEEE, 2011, 5(5):912-926.
- [47] Wipf D P, Rao B D, P. Wipf D, et al. Sparse Bayesian Learning for Basis Selection [J]. IEEE T. Signal. Proces. ,2004,52(8):2153-2164.
- [48] Donoho D L, Elad M. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization[J]. Proceedings of the National Academy of Sciences of the United States of America, National Acad Sciences, 2003, 100(5):2197.
- [49] Wipf D. Sparse Estimation with Structured Dictionaries [C] //NIPS. 2011(2):1-9.
- [50] Wipf D, Nagarajan S. Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions [J]. IEEE Journal of Selected Topics in Signal Processing, 2010, 4(2):317-329.
- [51] Candes E J, Wakin M B, Boyd S P. Enhancing sparsity by reweighted ℓ_1 minimization [J]. J Fourier Anal Appl, 2008, 14:877-905.
- [52] Wipf D, Rao B. Latent variable Bayesian models for promoting sparsity [J]. Information Theory, IEEE Transactions on, 2011, 57(9):6236-6255.
- [53] Zhang Z, Jung T-P, Makeig S, et al. Low energy wireless body-area networks for fetal ECG telemonitoring via the framework of block sparse Bayesian learning [J]. Arxiv preprint arXiv:1205.1287v1, 2012.
- [54] Wan J, Zhang Z, Yan J, et al. Sparse Bayesian Multi-Task

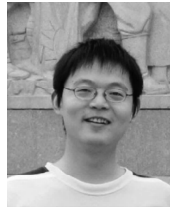
- Learning for Predicting Cognitive Outcomes from Neuroimaging Measures in Alzheimers Disease [C] // CVPR. 2012.
- [55] Zhang Z, Jung T-P, Makeig S, et al. Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware [J]. accepted by IEEE Trans. on Biomedical Engineering, 2012.
- [56] Makeig S, Kothe C, Mullen T, et al. Evolving signal processing for brain-computer interfaces [J]. Proceedings of the IEEE, 2012, 100: 1567-1584.
- [57] Neal R M. Bayesian learning for neural networks [M]. Springer, 1996.
- [58] Babacan S D, Molina R, Katsaggelos A K. Bayesian Compressive Sensing Using Laplace Priors [J]. IEEE T. Image. Process., 2010, 19(1): 53-63.
- [59] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables [J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Wiley Online Library, 2006, 68(1): 49-67.
- [60] Zhang Z, Rao B D D. Extension of SBL Algorithms for the Recovery of Block Sparse Signals with Intra-Block Correlation [J]. Arxiv preprint arXiv:1201.0862, 2012.
- [61] Rao B D, Zhang Z, Jin Y. Sparse signal recovery in the presence of intra-vector and inter-vector correlation [C] // International Conference on Signal Processing and Communications. 2012.
- [62] Zhang Z, Rao B D. Exploiting Correlation in Sparse Signal Recovery Problems: Multiple Measurement Vectors, Block Sparsity, and Time-Varying Sparsity [C] // ICML 2011 Workshop on Structured Sparsity: Learning and Inference. 2011.
- [63] Zhang Z, Rao B D. Iterative reweighted algorithms for sparse signal recovery with temporally correlated source vectors [C] // Proc. of the 36th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011). Prague, the Czech Republic; 2011.
- [64] Tang G, Nehorai A. Performance Analysis for Sparse Support Recovery [J]. IEEE Transactions on Information Theory, 2010, 56(3): 1383-1399.
- [65] Wipf D P, Rao B D. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem [J]. IEEE Transactions on Signal Processing, 2007, 55(7): 3704-3716.
- [66] Chen B, Paisley J, Carin L. Sparse Linear Regression with Beta Process Priors [C] // Proc. ICASSP. Dallas, Texas, USA; 2010: 1234-1237.
- [67] Soussen C, Idier J, Brie D, et al. From Bernoulli-Gaussian deconvolution to sparse signal restoration [J]. Signal Processing, IEEE Transaction on, 2010, 59(10): 4572-4584.
- [68] Herzet C, Drémeau A. Bayesian Pursuit Algorithms [J]. submitted to IEEE Trans. on Signal Processing, 2012.
- [69] Yu L. Reconstruction du signal ou de l'état basé sur un espace de mesure de dimension réduite, Signal state Reconstruction from Lower-dimensional Measurements [D]. University of Cergy Pontoise, 2011: 164.
- [70] Thibaux R, Jordan M I. Hierarchical Beta Processes and the Indian Buffet Process [C] // International Conference on Artificial Intelligence and Statistics. 2007.
- [71] Faktor T, Eldar Y C, Elad M, et al. Exploiting Statistical Dependencies in Sparse Representations for Signal Recovery [J]. IEEE Trans. on Signal Processing, 2012, 60(5): 2286-2303.
- [72] Faktor T, Eldar Y C, Elad M. Denoising of image patches via sparse representations with learned statistical dependencies [C] // 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011: 5820-5823.
- [73] Eduardo C, Adrian P O, Pedro S. Implementation of compressed sensing in telecardiology sensor networks [J]. International Journal of Telemedicine and Applications, 2010.
- [74] Dixon A M R, Allstot E G, Gangopadhyay D, et al. Compressed sensing system considerations for ECG and EMG wireless biosensors [J]. IEEE Trans. on Biomedical Circuits and Systems, 2012, 6(2): 156-166.
- [75] Mamaghanian H, Khaled N, Atienza D, et al. Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes [J]. IEEE Trans. on Biomedical Engineering, 2011, 58(9): 2456-2466.
- [76] Calhoun B H, Lach J, Stankovic J, et al. Body Sensor Networks: A Holistic Approach From Silicon to Users [J]. Proceedings of the IEEE, 2012, 100(1): 91-106.
- [77] Milenkovic A, Otto C, Jovanov E. Wireless sensor networks for personal health monitoring: Issues and an implementation [J]. Computer communications, 2006, 29(13-14): 2521-2533.
- [78] Makeig S, Bell A J, Jung T P, et al. Independent Component Analysis of Electroencephalographic Data [C] // Advances in neural information processing systems. 1996: 145-151.

- [79] Needell D, Tropp J A. CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples [J]. *Applied and Computational Harmonic Analysis*, 2009, 26 (3): 301-321.
- [80] Zou H, Hastie T. Regularization and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society: Series B*, 2005, 67(2): 301-320.
- [81] Mohimani H, Babaie-Zadeh M, Jutten C. A fast approach for overcomplete sparse decomposition based on smoothed L0 norm [J]. *IEEE Trans. on Signal Processing*, 2009, 57 (1): 289-301.
- [82] Huang J, Zhang T, Metaxas D. Learning with Structured Sparsity [C] // ICML. Montreal, Quebec, Canada; 2009.
- [83] Candès E J, Romberg J, Tao T. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information [J]. *IEEE T. Inform. Theory.*, 2006, 52(2): 489-509.
- [84] Yu L, Barbot J-P, Zheng G, et al. Compressive Sensing for Cluster Structured Sparse Signals: Variational Bayes Approach [R]. 2010.
- [85] Chen S S, Donoho D L, Saunders M A. Atomic Decomposition by Basis Pursuit [J]. *SIAM Journal on Scientific Computing, Society for Industrial and Applied Mathematics*, 1998, 20(1): 33-61.

作者简介



孙洪(1954-),出生于武汉。华中理工大学(现名:华中科技大学)获得工学博士学位,现为武汉大学电子信息学院教授,博士生导师。主要研究方向为统计信号处理、自适应信号处理和智能信号处理。E-mail: hongsun@whu.edu.cn



张智林(1980-),男,生于湖南,现于加州大学圣地亚哥分校(University of California, San Diego)电子与计算机工程系攻读博士学位,2002年毕业于电子科技大学获学士学位。主要研究方向为:稀疏贝叶斯学习,稀疏信号恢复,盲信号分离,医学信号处理,脑-机接口。E-mail: z4zhang@ucsd.edu



余磊(1985-),男,生于湖北钟祥,2006年毕业于武汉大学电子信息学院获学士学位,2012年毕业于武汉大学电子信息学院及法国国立高等应用电子学院(ENSEA)获得博士学位(联合培养),现于法国国立计算机及自动化研究院(INRIA-Bretagne Atlantique)博士后研究员。主要研究方向为:稀疏信号处理、压缩感知与医学图像处理。E-mail: lei.yu@inria.fr