

基于隐变量贝叶斯模型的稀疏信号恢复

王峰^{*①②} 向新^② 易克初^① 熊磊^②

^①(西安电子科技大学综合业务网国家重点实验室 西安 710071)

^②(空军工程大学航空航天工程学院 西安 710038)

摘要: 该文基于贝叶斯分析的视角, 揭示了一类算法, 包括使用隐变量模型的稀疏贝叶斯学习(SBL), 正则化FOCUSS算法以及Log-Sum算法之间的内在关联。分析显示, 作为隐变量贝叶斯模型的一种, 稀疏贝叶斯学习使用第2类最大似然(Type II ML)在隐变量空间进行运算, 可以视作一种更为广义和灵活的方法, 并且为不适定反问题的稀疏求解提供了改进的途径。较之于目前基于第1类最大似然(Type I ML)的稀疏方法, 仿真实验证实了稀疏贝叶斯学习的优越性能。

关键词: 信号处理; 隐变量贝叶斯模型; 第2类最大似然; 稀疏贝叶斯学习; 迭代加权最小二乘法

中图分类号: TN911.72

文献标识码: A

文章编号: 1009-5896(2015)01-0097-06

DOI: 10.11999/JEIT140169

Sparse Signals Recovery Based on Latent Variable Bayesian Models

Wang Feng^{①②} Xiang Xin^② Yi Ke-chu^① Xiong Lei^②

^①(State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China)

^②(Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract: From a Bayesian perspective, the commonly used sparse recovery algorithms, including Sparse Bayesian Learning (SBL), Regularized FOCUSS (R_FOCUSS) and Log-Sum, are compared. The analysis shows that, as a special case of latent variable Bayesian models, SBL, which operates in latent variable space via type-II maximum likelihood method, can be viewed as a more general and flexible means, and offers an avenue for improvement when finding sparse solutions to underdetermined inverse problems. Numerical results demonstrate the superior performance of SBL as compared to state-of-the-art sparse methods based on type-I maximum likelihood.

Key words: Signal processing; Latent variable Bayesian models; Type II Maximum Likelihood (ML); Sparse Bayesian Learning (SBL); Iteratively Reweighted Least Squares (IRLS)

1 引言

压缩感知理论^[1]的基础思想之一, 是求解一个不适定反问题, 因为“信号稀疏”或者“可压缩”这一先验的存在, 使唯一稀疏解的获得成为可能。目前已有的恢复算法至少可以划分为5类, 本文以其中基于隐变量模型的稀疏贝叶斯学习^[2](SBL)为主要研究对象。基于贝叶斯分析的视角, 在SBL和现有若干文献的方法之间建立了关联, 并深入分析了隐变量模型在信号恢复中的优势。

自2001年正式被提出后, SBL获得了广泛关注。首先, 文献[3]基于快速算法, 提出贝叶斯压缩感知(BCS), 随后, 文献[4]将稀疏贝叶斯学习理论引入到压缩感知领域, 并从变分角度进行了深入研究。2010年, 文献[5]巧妙地利用了一个3层概率模

型, 将Laplace先验引入贝叶斯压缩感知, 并证实了该模型是一个更为广义的表达。在算法性能提升上, 文献[6,7]都使用了BP (Belief Propagation)方法来寻求高效计算, 但二者使用的模型有所差异。此外, 文献[8]通过引入一个相关矩阵 B , 研究了向量之间存在相关特性时的贝叶斯恢复方法, 这种对先验信息的使用极大地提高了SBL的性能并扩展了应用范围。最近, 文献[9]也推导出了一类性能优异的算法。而文献[10]在结构噪声背景下提出了统计门限来增强SBL的计算性能。在理论性能研究方面, 有研究者推导了SBL的Cramér-Rao界, 提供了一个很有用的理论限^[11]。

在应用领域, 2001年就有研究者将SBL应用于信道均衡^[12], 后来被扩展到基于CMA结构的盲均衡器^[13]与稀疏多径信道的估计^[14]。在2013年权威期刊检索可以发现, SBL正被研究者扩展至无线信道建模^[15]和电力线通信系统中的冲激噪声抑制^[16]等领

2014-01-24收到, 2014-06-13改回

国家自然科学基金(61379104)资助课题

*通信作者: 王峰 wangfengsn@163.com

域。此外,国内的研究者也将SBL理论应用于DOA估计等问题的实现^[17]。

基于隐变量贝叶斯模型的方法已经获得了广泛的关注,但有一些问题并没有得到明确系统的回答,首先,稀疏贝叶斯学习和已有算法之间有什么内在关联?其次,其稀疏解的获得和解的优越性从何而来?这些问题在本文中得到了部分解答。

论文后续的部分组织结构如下。第2节给出了SBL信号模型和ARD(Automatic Relevance Determination)先验假设^[18];第3节介绍了第1类最大似然和使用隐变量模型的第2类最大似然方法,指出SBL在稀疏信号恢复中的优势,特别是对局部解的平滑;第4节揭示了SBL与现有文献算法的内在关联;第5节给出了仿真结果;第6节总结全文。

2 信号模型

系统存在噪声时,压缩感知的标准模型表示为 $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$,其中 $\mathbf{x} \in R^N$, $\mathbf{y} \in R^M$,且 $\mathbf{A} \in R^{M \times N}$ 是一个满足RIP准则的测量矩阵, \mathbf{n} 表示在测量中所引入的噪声,一般假设 $\mathbf{n} \sim N(0, \sigma^2)$ 。因为 $N > M$,该问题通常没有唯一解,但如果为 \mathbf{x} 施加一个稀疏约束,或者使用稀疏诱导先验,则可以得到唯一最稀疏解。因为 l_0 范数的最小化是一个NP-难问题,所以一般使用较为松弛的先验。

假设 $\mathbf{x} \sim p(\mathbf{x})$,使用最大后验概率(Maximum A Posteriori, MAP)准则,则寻求最佳解的问题就转化为

$$\begin{aligned} \mathbf{x}_{\text{MAP}} &= \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}) \\ &= \arg \min_{\mathbf{x}} [-\ln p(\mathbf{y} | \mathbf{x}) - \ln p(\mathbf{x})] \\ &= \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 - \lambda \ln p(\mathbf{x}) \end{aligned} \quad (1)$$

对照Tikhonov正则化表达式,可以发现 $\lambda \ln p(\mathbf{x})$ 项起到了正则化的作用,在文献中也称之为稀疏诱导或者稀疏惩罚函数。其中 λ 是一个调节参数,用于在恢复信号的稀疏度和均方意义下的精度之间进行折中,来实现有偏的稀疏估计。对稀疏约束先验 $p(\mathbf{x})$ 的选择具有很大的灵活性,那些在集中在0周围且具有重拖尾的概率密度函数能够满足要求,比如可以使用广义高斯分布(Generalized Gaussian Density, GGD)来表示某一族函数。

$$f(u; p, \beta) = \frac{p}{2\sqrt{2}\beta\Gamma(1/p)} e^{-(|u|^p/2\beta^p)}, \quad p > 0$$

当参数 β 和 p 取不同值的时候,就会得到不同的先验和算法,如岭回归, LASSO等。

SBL所使用的ARD先验,使用了独立高斯先验假设,即假设 $p(\mathbf{x}; \boldsymbol{\gamma}) = N(\mathbf{x}; 0, \text{diag}[\boldsymbol{\gamma}])$,考虑到噪

声的存在时,又有 $p(\mathbf{y} | \mathbf{x}) = N(\mathbf{y}; \mathbf{A}\mathbf{x}, \sigma^2 \mathbf{I})$ 。通常情况下,高斯先验不能认为是一种稀疏约束,于是使用分层先验的思想,为每个 γ_i 设定了超分布:

$$p(\gamma_i^{-1}) = \Gamma(a)^{-1} b^a \gamma_i^{1-a} e^{(-b/\gamma_i)}$$

这是一个典型的多层先验模型。ARD模型中的 $\boldsymbol{\gamma}$ 和 σ^2 (见下文)称之为超参数。

对超参数使用Gamma先验的原因是为了获得解析解,其中 a 表示形状参数, b 表示尺度参数。在ARD先验中假设 $p(\sigma^{-2}) = \text{Gamma}(\sigma^{-2} | c, d)$,但同时 c, d 趋于0,且该先验不传递,故作一般变量处理。需要指出的是,文献[2,4]分别使用 γ_i 和 $\alpha_i = 1/\gamma_i$ 作为超参数,仅是表达上的不同。

3 第1类与第2类最大似然

3.1 第1类最大似然

基于ARD先验假设,直接使用式(1)的MAP准则,令 $p(\boldsymbol{\gamma})$ 中 $a, b \rightarrow 0$,会得到如下的目标函数:

$$\begin{aligned} L_1^x(\mathbf{x}, \boldsymbol{\gamma}) &= \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \left(\mathbf{x}^T \boldsymbol{\Gamma}^{-1} \mathbf{x} + \sum_{i=1}^N \ln |\gamma_i| \right) \end{aligned} \quad (2)$$

式(2)是一个二元优化问题,可使用坐标下降法迭代求解,其中 $\lambda = \sigma^2$, $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$, $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_N\}$ 。

此外,还有一种分析的思路,如果通过边际积分消掉 $\boldsymbol{\gamma}$,也即将超参数 $\boldsymbol{\gamma}$ 的先验“传递”给 \mathbf{x} ,就得到如下结论:

$$\begin{aligned} p(x_i) &= \int p(x_i | \gamma_i) p(\gamma_i) d\gamma_i \\ &= \frac{b^a \Gamma(a + 1/2)}{(2\pi)^{1/2} \Gamma(a)} \left(b + \frac{x_i^2}{2} \right)^{-(a+1/2)} \\ &\xrightarrow{a, b \rightarrow 0} \propto \frac{1}{|x_i|} \end{aligned}$$

此处令 $a, b \rightarrow 0$,是为了获得更为稀疏的先验。同样使用式(1),得到目标函数:

$$L(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^N \ln(|x_i|)$$

由此可以看出,使用第1类最大似然并边际积分后,得到一个正则化的Log-Sum最小化目标函数,显然该目标函数是非凸的。

Log-Sum最小化的问题定义如下:

$$\min_{\mathbf{x}} L(\mathbf{x}) = \sum_{i=1}^N \ln(|x_i| + \varepsilon), \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{y}$$

其中 ε 是一个极小的值用以避免出现负无穷的目标值。

3.2 第2类最大似然

第2类最大似然也被称为经验贝叶斯,和第1类最大似然的信号模型相同,但是求解途径不同。

前者将 \mathbf{x} 视作隐变量，首先将稀疏向量问题视作高斯模型选择问题，模型选择成功后，再使用式(1)求解向量。

引入超参数 γ 和 σ^2 后，问题转化为

$$\begin{aligned} (\hat{\mathbf{x}}, \hat{\gamma}, \hat{\sigma}^2) &= \arg \max_{\mathbf{x}, \gamma, \sigma^2} p(\mathbf{x}, \gamma, \sigma^2 | \mathbf{y}) \\ &= \arg \max_{\mathbf{x}, \gamma, \sigma^2} p(\mathbf{x} | \mathbf{y}, \gamma, \sigma^2) p(\gamma, \sigma^2 | \mathbf{y}) \end{aligned} \quad (3)$$

式(3)的第1项表示两个超参数 γ 与 σ^2 已知时 \mathbf{x} 的分布，使用前面的高斯假设，可以求得 $p(\mathbf{x} | \mathbf{y}; \gamma, \sigma^2) = N(\mu, \Sigma_x)$ ，其中 $\mu = \sigma^{-2} \Sigma_x \mathbf{A}^T \mathbf{y}$ ， $\Sigma_x = (\sigma^{-2} \mathbf{A}^T \mathbf{A} + \Gamma^{-1})^{-1}$ ，且 $\Gamma = \text{diag}(\gamma)$ 。

式(3)的第2项可以求得 $p(\mathbf{y}; \gamma, \sigma^2) = N(0, \Sigma_y)$ ，其中 $\Sigma_y = \sigma^2 \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T$ 。

寻求式(3)的解分可以为两步，首先寻求使得 $p(\mathbf{y}; \gamma, \sigma^2)$ 最大的超参数(通过边际积分消掉 \mathbf{x} ，即最大边际似然)。然后，利用所求的当前超参数和关系式 $p(\mathbf{x} | \mathbf{y}; \gamma, \sigma^2)$ ，基于 MAP 准则，寻求 \mathbf{x} 的最佳解。不难看出，这个运算流程构成了迭代过程，通过数次迭代后， \mathbf{x} 大部分中大部分元素被置 0，余下元素可以认为是相关向量，这就是相关向量机的由来。限于篇幅，解过程和形式在此不再详述，见文献[2, 4]。

3.3 两类最大似然方法的关系

可以看出，第1类与第2类最大似然使用了相同的信号模型，但计算过程迥异，前者的计算主要体现在 \mathbf{x} 空间，后者因为将 \mathbf{x} 视作隐变量在边缘积分中消掉，所以主要在 γ 空间分析。文献[19]给出了两类最大似然在形式上的等效表达，本文进一步发展了该表达，并得到一些有益的分析结论。

定理 1 ARD 先验的假设条件下，在 \mathbf{x} 空间或者 γ 空间观察，使用第1类最大似然所取得的目标函数与使用第2类最大似然所取得目标函数均相差一个因子：

$$\Delta_{1,2} = \ln |\lambda^{-1} \mathbf{A}^T \mathbf{A} + \Gamma^{-1}|$$

其中 $\lambda = \sigma^2$ 。

证明 (1)从 \mathbf{x} 空间分析。如前所述，基于 MAP 准则，使用第1类最大似然的目标函数式(2)可以等效表达为

$$L_{(1)}^x(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \left(\sum_{i=1}^N \frac{x_i^2}{\gamma_i} + \sum_{i=1}^N \ln(\gamma_i) \right) \quad (4)$$

将第2类最大似然的目标函数转化到 \mathbf{x} 空间，其可以表达为^[19]

$$\begin{aligned} L_{(2)}^x(\mathbf{x}) &= \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \left(\sum_{i=1}^N \frac{x_i^2}{\gamma_i} + \ln |\lambda \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T| \right) \\ &= \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \\ &\quad + \lambda \left(\sum_{i=1}^N \frac{x_i^2}{\gamma_i} + \ln |\Gamma| + \ln |\lambda^{-1} \mathbf{A}^T \mathbf{A} + \Gamma^{-1}| \right) \end{aligned}$$

上式使用了关系式：

$$\ln |\lambda \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T| = \ln |\Gamma| + \ln |\lambda^{-1} \mathbf{A}^T \mathbf{A} + \Gamma^{-1}| \quad (5)$$

根据 $\Gamma = \text{diag}(\gamma)$ ，可以证明在 \mathbf{x} 空间的结论。

(2)从 γ 空间分析。首先将第1类最大似然函数从 \mathbf{x} 空间转换至 γ 空间，将式(4)除以 λ ，并不影响其优化解，使用关系式：

$$\begin{aligned} \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} &= \mathbf{y}^T (\sigma^2 \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T)^{-1} \mathbf{y} \\ &= \sigma^{-2} \mathbf{y}^T \mathbf{y} - \sigma^{-2} \mathbf{y}^T \mathbf{A} (\sigma^{-2} \mathbf{A}^T \mathbf{A} + \Gamma^{-1})^{-1} \mathbf{A}^T \sigma^{-2} \mathbf{y} \\ &= \sigma^{-2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \sigma^{-2} \mathbf{x}^T \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{x}) \\ &= \sigma^{-2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \mathbf{x}^T \Gamma^{-1} \mathbf{x} \\ &= \lambda^{-1} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \mathbf{x}^T \Gamma^{-1} \mathbf{x} \end{aligned}$$

得到第1类最大似然的目标函数：

$$L_{(1)}^{\gamma}(\gamma) = \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} + \sum_{i=1}^N \ln(\gamma_i) \quad (6)$$

此外，第2类最大似然的目标函数可以表示为^[4]：

$$L_{(2)}^{\gamma}(\gamma) = \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} + \ln |\Sigma_y| \quad (7)$$

再次使用式(5)，可以得到在 γ 空间的结论。证毕

定理 1 说明，两类最大似然方法是求解同一个问题的不同途径，二者因为计算路径的不同带来了目标函数上的差异，进一步分析和仿真发现，这个差异因子 $\Delta_{1,2}$ 使得 SBL 获得了平滑局部解的能力。

关于 SBL 对局部解的平滑问题，目前并没有十分清晰的解释，一个明确的结论是，第2类最大似然方法所得到的稀疏诱导函数是不可分的，正是分量之间的互相制约，起到了平滑效果，感兴趣的读者可参考文献[19]，此外，对该问题一个直观且可行的解释是从变分逼近的观点出发，在文献[20]中有详细描述。

4 与现有算法的关系

4.1 R-FOCUSS 与第1类最大似然

使用坐标下降法求解式(2)，即分别令 $\partial L(\mathbf{x}, \gamma) / \partial \gamma_i = 0$ 和 $\partial L(\mathbf{x}, \gamma) / \partial x_i = 0$ 分别可以求得

$$\left. \begin{aligned} \gamma_i &= x_i^2 \\ \mathbf{x} &= \Gamma \mathbf{A}^T (\lambda \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T)^{-1} \mathbf{y} \end{aligned} \right\}$$

对照文献[21]，此即正则化的 FOCUSS 方法，简称 R-FOCUSS。

4.2 极限情况

由恒等式 $(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1} \mathbf{A} \equiv \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}$ ，可以证明 $\Gamma \mathbf{A}^T (\lambda \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T)^{-1} = (\mathbf{A}^T \mathbf{A} + \lambda \Gamma^{-1})^{-1} \mathbf{A}^T$ ，设 $\Gamma = \mathbf{W}^{(k)}$ ，R-FOCUSS 算法可以表示为

$$\begin{aligned}
\hat{\mathbf{x}} &= \mathbf{\Gamma} \mathbf{A}^T (\lambda \mathbf{I} + \mathbf{A} \mathbf{\Gamma} \mathbf{A}^T)^{-1} \mathbf{y} \\
&= \mathbf{\Gamma}^{1/2} (\mathbf{A} \mathbf{\Gamma}^{1/2})^T (\lambda \mathbf{I} + (\mathbf{A} \mathbf{\Gamma}^{1/2}) (\mathbf{A} \mathbf{\Gamma}^{1/2})^T)^{-1} \mathbf{y} \\
&= \lim_{\lambda \rightarrow 0} \mathbf{\Gamma}^{1/2} (\mathbf{A} \mathbf{\Gamma}^{1/2})^T (\lambda \mathbf{I} + (\mathbf{A} \mathbf{\Gamma}^{1/2}) (\mathbf{A} \mathbf{\Gamma}^{1/2})^T)^{-1} \mathbf{y} \\
&= \mathbf{\Gamma}^{1/2} (\mathbf{A} \mathbf{\Gamma}^{1/2})^+ \mathbf{y} \tag{8}
\end{aligned}$$

式(8)就是问题 $L_{\text{FOCUSS}} = \|\mathbf{\Gamma} \mathbf{x}\|_2^2$, s.t. $\mathbf{y} = \mathbf{A} \mathbf{x}$ 的解。

因为 SBL 具有和 R-FOCUSS 类似的解结构(二者都是高斯先验假设, 差异在于权值即先验方差的获取方式不同), 所以当噪声为 0 的极限情况下, SBL 也会收敛到 FOCUSS, 这一点和文献[4]的结论是完全一致的。

根据这种相似性, 我们实际上是为 R-FOCUSS 这一类正则化迭代加权最小二乘法 (Iteratively Reweighted Least Squares, IRLS) 算法的反馈权值赋予了物理意义——权值即是超参数先验模型中的标准差。

更进一步地, 通过对现有文献的分析, 有如下的结论: (1)FOCUSS 和 IRL1 的每一步求解, 都相当于在逼近一个 Log-Sum 问题^[21, 22]; (2)当 $\varepsilon \approx 0$ 时, Log-Sum 和 $\min l_0$ 是等效的^[23]; (3)SBL 和 FOCUSS 都是正则化的 $\min l_0$ 问题; (4)SBL 求解的也是 Log-sum 方法(见本文第 3 节)。该结论可以总结如图 1 所示。

从图 1 中可以看出, SBL 是一种广义的、引入了更多参数的正则化 l_0 范数最小化计算方法, 所以在获得稀疏提升的同时, 具备了很强的灵活性, 当然随着也带来了计算上的复杂度。

5 数值仿真与分析

5.1 SBL 在局部解方面的优势

问题设定如下, 假设某一向量 \mathbf{x}^{11} , 且 $|\mathbf{x}|_0 = 10$, 存在某一矩阵 $\mathbf{A} = \text{randn}(10, 11)$, 假设矩阵零空间向量 $\hat{\mathbf{v}} \in \text{null}(\mathbf{A})$, 则可以用 $\mathbf{v} = \mathbf{x} + \alpha * \hat{\mathbf{v}}$ 来表示反问题 $\mathbf{y} = \mathbf{A} \mathbf{x}$ 的解, 其中 $\alpha \in [-10 : 0.1 : 10]$ 表示尺度因子变量。假设目标函数为

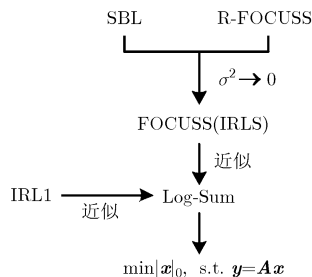


图 1 算法之间的关系示意图

$$L_p(\mathbf{v}) = \left(\sum_{i=1}^N |v_i|^p \right)^{1/p}$$

图 2 中给出了 $p=0.01$, $p=1.00$, 以及 Log-Sum 3 种情况下的目标函数示意图。

从图 2 中可以看出, 以 l_p 范数最小化为目标函数时, 当 $p=0.01 \approx 0$ 时, 因为目标函数的高度非凸性, 所以出现了很多局部解, Log-Sum 目标函数同样存在这种情况。但也可以看出, 此时目标函数的全局最小与最稀疏解是完全对应的, 也即保证了了解的稳定性, 这一点在直观上也容易理解, 因为局部解实际上发生在基本可行解处, 代表了目标函数对 0 值的敏感程度, 对局部解越敏感, 则全局解的稳定性就越高。

当 $p=1$ 时, 目标函数具有凸性, 所以不存在局部解, 但是因为过度光滑易导致解不稳定, 即目标函数的全局最小(对凸函数来说也是局部最小)没有发生在最稀疏即 $\alpha=0$ 处, 这种解的不稳定性, 就是 l_0 范数向 l_1 范数松弛所付出的代价。

在图 3 中给出了 SBL(第 2 类最大似然)和 Log-Sum(第 1 类最大似然)在局部解方面的对比, 此时 $|\mathbf{x}|_0 = 7$, $\alpha \in [-15 : 0.1 : 15]$, 其他条件不变。从图中可以看出, SBL 的局部解比 Log-Sum 要少, 前者为 4 个而后者有 6 个, 而且又保持了稀疏解的稳定性。事实上如果 $|\mathbf{x}|_0$ 足够小, \mathbf{A} 选择合适的话, SBL 可以保持无局部解。

根据仿真结果, SBL 中局部解的数目和矩阵的相干特性、信号结构以及信号的稀疏度都有关联, 文献[20]对其进行了初步分析。

5.2 相对于其他算法的恢复性能优势

本部分使用两种 SBL 的实现算法, 分别是贝叶斯压缩感知 (Bayesian Compressive Sensing, BCS), EM-SBL, 与其他的贪婪算法或者凸松弛算法进行了仿真比较, 这些算法包括: CoSaMP, SP (Subspace Pursuit), FOCUSS, l_1 -Magic。

仿真条件设定如下: $\mathbf{A} = \text{randn}(80, 390)$, 信号 \mathbf{x} 的稀疏度 $k=|\mathbf{x}|_0=15$ 。对于 CoSaMP 和 SP 算法, 输入参数包括非零元素的个数, 方便其判断收敛和提高性能。对于 l_1 -Magic 和 FOCUSS 算法, 提供真实的信噪比, 以便计算相关调节参数。每种算法迭代 300 次, 统计其正确恢复概率。

信噪比定义为

$$\text{SNR}(\text{dB}) = 20 \lg \left(\|\mathbf{A} \mathbf{x}\|^2 / \|\mathbf{n}\|^2 \right)$$

仿真度量指标为恢复成功概率, 这里所谓的恢复成功, 指的是信号 \mathbf{x} 的支撑与恢复信号 $\hat{\mathbf{x}}$ 的前 k 个较大幅值元素的支撑相同。

从图 4 可以看出, EM-SBL 因为其基于 EM 算法计算, 所以性能最优, 在信噪比较小的时候即可实现该概率信号支撑恢复。而 BCS 因为使用了快速算法, 所以性能下降比较严重。但是基于 SBL 方法的一个显著优势是参数可以自适应计算, 无需提前指定。

5.3 算法复杂性分析

FOCUSS 算法的运算负担来自于求解第 k 次运算结果 $\hat{x}^{(k)}$ 时的矩阵求逆 (k 表示迭代次数):

$$\hat{x}^{(k)} = \mathbf{W}^* \left((\mathbf{A} \mathbf{W}^*)^T (\mathbf{A} \mathbf{W}^*) + \lambda \mathbf{I} \right)^{-1} (\mathbf{A} \mathbf{W}^*)^T \mathbf{y}$$

其中 $\mathbf{W}^* = \mathbf{W}_{(k-1)}$ 为第 $k-1$ 次所产生的反馈权值。EM-SBL 涉及到两次矩阵求逆, 分别是求取 $\hat{\mathbf{x}}^{(k)} = (\mathbf{A}^T \mathbf{A} + \sigma^2 \Gamma_{(k)}^{-1})^{-1} \mathbf{A}^T \mathbf{y}$ 和求取第 k 次超参数时的 $\gamma_{i,(k)} = (\sigma^2 \mathbf{A}^T \mathbf{A} + \Gamma_{(k-1)}^{-1})^{-1}_{i,i} + x_{i,(k-1)}^2$ 。在经过快速算法处理之后, 所获得的 BCS 进行了简化处理, 运

算量有所降低, 基本与 FOCUSS 相当。SP 算法和 CoSaMP 都是用了最小二乘, 但是维数不同, 所以运算速度上前者较快。本文使用 l_1 -Magic 中的障碍函数法求解, 该方法是基于牛顿法实现的, 也具有运算上的优势。

因为各类算法达到收敛的迭代次数和运算结果差异, 所以本文采用固定运算次数的方法来来进行比较。使用 ThinkPad T420 笔记本电脑, Winxp 系统, 2G 内存, 仿真平台为 MatlabR2011a, 每种算法各仿真 100 次。结果如下(单位为 s): CoSaMP(内部固定循环 100 次): 13.142959; BCS: 11.307114; EM-SBL: 63.925165; l_1 -Magic: 7.423480; FOCUSS: 17.372597; SP(内部固定循环 100 次): 3.318045。从仿真结果可以看出, SBL 类算法的运算量还有一定的优化余地。

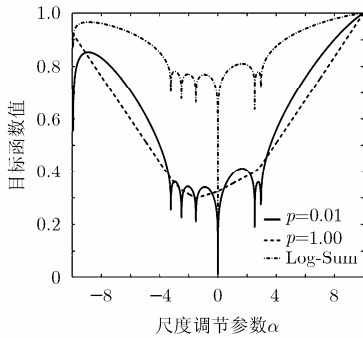


图 2 各种稀疏先验下解的稳定性

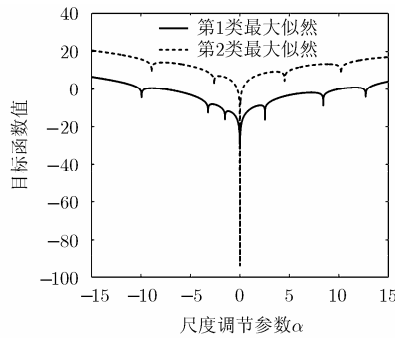


图 3 SBL 对局部解的平滑

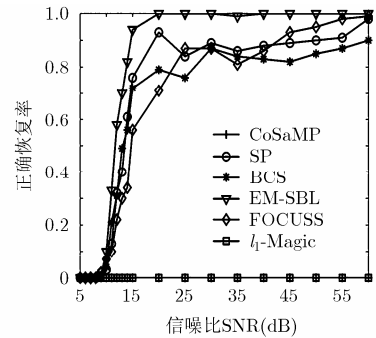


图 4 SBL 与其他算法恢复性能对比图

6 结束语

本文使用贝叶斯分析框架, 对基于隐变量模型的稀疏贝叶斯学习方法进行了研究分析, 得出结论如下: (1)使用 ARD 先验的第 1 类和第 2 类最大似然之间有密切的关联, 二者相差一个参数因子。(2)SBL 与 FOCUSS, Log-Sum 以及 $\min l_0$ 问题存在着内在关联。(3)稀疏贝叶斯学习在保持全局解稳定的同时, 部分局部解得到了平滑, 且其在噪声环境下的恢复性能较为优异。

SBL 的一个局限之处在于计算复杂度, 进一步的研究, 可以考虑发展新的快速算法, 在复杂度和性能之间寻求最佳结合点。

参考文献

[1] Candès E J and Wakin M B. An introduction to compressive sampling[J]. *IEEE Signal Processing Magazine*, 2008, 25(2): 21-30.
 [2] Tipping M E. Sparse Bayesian learning and the relevance

vector machine[J]. *Journal of Machine Learning Research*, 2001, 1(6): 211-244.

[3] Ji S, Xue Y, and Carin L. Bayesian compressive sensing[J]. *IEEE Transactions on Signal Processing*, 2008, 56(6): 2346-2356.
 [4] Wipf D P and Rao B D. Sparse Bayesian learning for basis selection[J]. *IEEE Transactions on Signal Processing*, 2004, 52(8): 4036-4048.
 [5] Babacan S D, Molina R, and Katsaggelos A K. Bayesian compressive sensing using Laplace priors[J]. *IEEE Transactions on Image Processing*, 2010, 19(1): 53-63.
 [6] Tan X and Li J. Computationally efficient sparse Bayesian learning via belief propagation[J]. *IEEE Transactions on Signal Processing*, 2010, 58(4): 2010-2021.
 [7] Baron D, Sarvotham S, and Baraniuk R. Bayesian compressive sensing via belief propagation[J]. *IEEE Transactions on Signal Processing*, 2010, 58(1): 269-280.
 [8] Zhang Z and Rao B D. Extension of SBL algorithms for the

- recovery of block sparse signals with intra-block correlation[J]. *IEEE Transactions on Signal Processing*, 2013, 61(8): 2009–2015.
- [9] Qui K and Dogandzic A. Sparse signal reconstruction via ECME hard thresholding[J]. *IEEE Transactions on Signal Processing*, 2012, 60(9): 4551–4569.
- [10] Hurtado M, Muravchik C H, and Nehorai A. Enhanced sparse Bayesian learning via statistical thresholding for signals in structured noise[J]. *IEEE Transactions on Signal Processing*, 2013, 61(21): 5430–5443.
- [11] Prasad R and Murthy C R. Cramér-Rao-Type bounds for sparse Bayesian learning[J]. *IEEE Transactions on Signal Processing*, 2013, 61(3): 622–632.
- [12] Chen S, Gunn S R, and Harris C J. The relevance vector machine technique for channel equalization application[J]. *IEEE Transactions on Neural Networks*, 2001, 12(6): 1529–1532.
- [13] Hwang K and Choi S. Blind equalization method based on sparse bayesian learning[J]. *IEEE Signal Processing Letters*, 2009, 16(4): 315–318.
- [14] Shutin D and Fleury B. Sparse variational Bayesian SAGE algorithm with application to the estimation of multipath wireless channels[J]. *IEEE Transactions on Signal Processing*, 2011, 59(8): 3609–3623.
- [15] Cal-Braz J A, Matos L J, and Cataldo E. The relevance vector machine applied to the modeling of wireless channels [J]. *IEEE Transactions on Antennas and Propagation*, 2013, 61(12): 6157–6167.
- [16] Lin J, Nassar M, and Evans B L. Impulsive noise mitigation in powerline communications using sparse bayesian learning[J]. *IEEE Journal on Selected Areas in Communications*, 2013, 31(7): 1172–1183.
- [17] 孙磊, 王华力, 许广杰, 等. 基于稀疏贝叶斯学习的高效 DOA 估计方法[J]. *电子与信息学报*, 2013, 35(5): 1196–1201.
- Sun Lei, Wang Hua-li, Xu Guang-jie, *et al.* Efficient direction-of-arrival estimation via sparse Bayesian learning[J]. *Journal of Electronics & Information Technology*, 2013, 35(5): 1196–1201.
- [18] Neal R. Bayesian learning for neural networks[D]. [Ph.D. dissertation], University of Toronto, 1996.
- [19] Wipf D P, Rao B D, and Nagarajan S. Latent variable Bayesian models for promoting sparsity[J]. *IEEE Transactions on Information Theory*, 2011, 57(9): 6236–6255.
- [20] Wipf D P. Bayesian methods for finding sparse representations[D]. [Ph.D. dissertation], University of California, San Diego, 2006.
- [21] Gorodnitsky I F and Rao B D. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm[J]. *IEEE Transactions on Signal Processing*, 1997, 45(3): 600–616.
- [22] Candès E J, Wakin M B, and Boyd S P. Enhancing sparsity by reweighted L1 minimization[J]. *Journal of Fourier Analysis and Applications*, 2008, 14(5): 877–905.
- [23] Shen Y, Fang J, and Li H. Exact reconstruction analysis of Log-Sum minimization for compressed sensing[J]. *IEEE Signal Processing Letters*, 2013, 20(12): 1223–1226.
- 王 峰: 男, 1978 年生, 博士生, 讲师, 研究方向为稀疏贝叶斯学习、基于压缩感知的通信信号处理.
- 向 新: 男, 1971 年生, 博士, 教授, 研究方向为通信信号处理、超宽带通信与综合航电系统.
- 易克初: 男, 1943 年生, 教授, 博士生导师, 研究方向为无线通信、扩频通信.
- 熊 磊: 男, 1976 年生, 博士, 副教授, 研究方向为模式识别、机器学习.