

参数自适应调整的稀疏贝叶斯重构算法

夏建明* 杨俊安 陈 功

(电子工程学院 合肥 230037)

(安徽省电子制约技术重点实验室 合肥 230037)

摘 要: 稀疏表示模型中的正则化参数由未知的噪声和稀疏度共同决定, 该参数的设置直接影响稀疏重构性能的好坏。然而目前稀疏表示问题优化求解算法或依靠主观、或依靠相关先验信息、或经过实验设置该参数, 均无法自适应地设置调整该参数。针对这一问题, 该文提出一种无需先验信息的参数自动调整的稀疏贝叶斯学习算法。首先对模型中各参数进行概率建模, 然后在贝叶斯学习的框架下将参数设置及稀疏求解问题转化为一组混合 L1 范数与加权 L2 范数之和的凸优化问题, 最终通过迭代优化得到参数设置和问题求解。由理论推导和仿真实验可知, 已知理想参数时, 该算法与其它非自动设置参数的迭代重加权算法性能相当, 甚至更优; 在理想参数未知时, 该算法的重构性能要明显优于其它算法。

关键词: 压缩感知; 稀疏重构; 迭代重加权; 稀疏贝叶斯学习; 参数自动调整

中图分类号: TN911.72

文献标识码: A

文章编号: 1009-5896(2014)06-1355-07

DOI: 10.3724/SP.J.1146.2013.00629

Bayesian Sparse Reconstruction with Adaptive Parameters Adjustment

Xia Jian-ming Yang Jun-an Chen Gong

(Electronic Engineering Institute, Hefei 230037, China)

(Key Laboratory of Electronic Restriction, Anhui Province, Hefei 230037, China)

Abstract: The regularization parameter of sparse representation model is determined by the unknown noise and sparsity. Meanwhile, it can directly affect the performances of sparsity reconstruction. However, the optimization algorithm of sparsity representation issue, which is solved with parameter setting by expert reasoning, priori knowledge or experiments, can not set the parameter adaptively. In order to solve the issue, the sparsity Bayesian learning algorithm which can set the parameter adaptively without priori knowledge is proposed. Firstly, the parameters in the model is constructed with the probability. Secondly, on the basis of the framework of Bayesian learning, the issue of parameter setting and sparsity resolving is transformed to the convex optimization issue which is the addition of a series of mixture L1 normal and the weighted L2 normal. Finally, the parameter setting and sparsity resolving are achieved by the iterative optimization. Theoretical analysis and simulations show that the proposed algorithm is competitive and even better compared with other parameter non-adjusted automatically iterative reweighted algorithms when ideal parameter is known, and the reconstruction performance of the proposed algorithm is significantly better than the other algorithms when choosing the non-ideal parameters.

Key words: Compressive sensing; Sparse signal reconstruction; Iterative reweighting; Sparse Bayesian learning; Parameters adjust automatically

1 引言

近几年来, 基于稀疏表示的机器学习方法在各领域得到迅速发展。将稀疏表示的思想引入解决机器学习的问题, 最初是由压缩感知、稀疏表示在信号处理中的应用启发而来。两个问题可抽象为统一的稀疏重构问题来进行描述。稀疏重构任务可归结

为寻找欠定线性系统 $y = \Phi x + n$ 的最稀疏解, 在求解过程中需要引入正则化参数来对数据拟合和信号稀疏度进行平衡。由于该参数一般为噪声和信号稀疏度的函数, 常常难以选择。而算法成功与否以及重构性能好坏对该参数的选择有很强的依赖性。现有的稀疏表示模型求解方法都无法自动设置、调整正则化参数。如凸松弛的方法, 例如基底追踪去噪 (Basis Pursuit DeNoising, BPDN)^[1], Dantzig 选择器 (Dantzig selector)^[2], 最小绝对值收缩选择器 (Least Absolute Shrinkage and Selection Operator,

2013-05-09 收到, 2014-01-15 改回

安徽省自然科学基金(1208085MF94, 1308085QF99)和国家自然科学基金(61272333) 资助课题

*通信作者: 夏建明 jianmingeei@163.com

LASSO)^[3]等。由于等效的优化目标函数为凸函数, 这些算法的主要共同优点是能够获得唯一解, 但大多数凸松弛算法均要求主观设置正则化参数; 另一类贪婪算法复杂度较低但重构性能不足, 也同样存在参数设置问题, 如: 正交匹配追踪(Orthogonal Matching Pursuit, OMP)^[4]、阶段正交匹配追踪(stagewise OMP)^[5]、子空间追踪(Subspace Pursuit)^[6]等。与上述两种逼近求解方法不同的另一种方法是稀疏贝叶斯学习方法(Sparse Bayesian Learning, SBL)。它基于贝叶斯思想, 对模型中未知量进行概率建模, 然后进行贝叶斯推理求得稀疏解^[7-11]。该模型在无噪声时的全局最优解与欠定线性系统的最稀疏解等价, 且在含噪声时所得的局部最优解也是稀疏的^[7]。文献[9]证明了其最差情况均要优于 LASSO, 且便于添加对稀疏信号的约束。但是, 贝叶斯稀疏重构算法假设噪声功率是已知的, 不能自适应进行调整^[12,13]。此后的许多迭代重加权算法均需事先设置或启发式设置^[14-16]参数。综上所述, 现有的优化算法都需要通过主观或是依据别的方法来设置正则化参数。

在优化算法框架之外, 对稀疏正则化参数数值选择方法的研究也陆续展开, 这些方法总的来说都是从 Tikhonov 正则化下^[17]的参数选择方法扩展而来, 主要包括 Discrepancy principle^[18], SURE (Stein's Unbiased Risk Estimator)^[19], GCV (Generalized Cross Validation)^[20], GSURE (Generalized SURE)^[21]和 L-curve^[22]等。但这些方法本质上需要噪声的先验知识, 或依靠反复实验来根据运算结果选取最佳参数。如 GCV 方法只能应用于高斯白噪声条件下; Discrepancy principle 和 SURE 需要预知噪声方差; GSURE 需要噪声分布参数; 而广泛使用的 L-curve 方法依赖于实验, 运算量巨大。

为了解决正则化参数自动设置、调整的问题, 本文推导了一种无需先验信息的参数自动调整的稀疏贝叶斯学习算法。通过迭代求解一些列混合 L2 范数与加权 L1 范数之和的凸优化问题来求解 SBL 问题。论文将在第 2 节针对上述问题构建贝叶斯稀疏重构模型; 接着在第 3 节推导参数自动调整的迭代重加权求解算法, 并在第 4 节对算法进行收敛性分析; 第 5 节将用数值仿真实验验证分析算法的重构性能; 最后在第 6 节进行讨论和小结。

2 贝叶斯稀疏重构模型

考虑线性数据生成模型 $\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}$, 其中 $F \in \mathbb{R}^{M \times N}$ 为已知的行满秩观测矩阵, $\mathbf{x} \in \mathbb{R}^N$ 为信号向量, $\mathbf{n} \in \mathbb{R}^M$ 为模型误差和观测噪声。贝叶斯模型即是将未知量建模为随机变量, 并赋予相应的概

率密度函数。若信号 $\mathbf{x}$ 和噪声功率 $\beta = \sigma^2$ 已知, 则观测数据 $\mathbf{y}$ 的概率密度函数建模为

$$p(\mathbf{y} | \Phi \mathbf{x}, \beta) = (2\pi\beta)^{-M/2} \exp \left\{ -\frac{1}{2} \frac{\|\mathbf{y} - \Phi \mathbf{x}\|^2}{\beta} \right\} \quad (1)$$

2.1 稀疏信号的概率模型构建

由于拉普拉斯先验分布在对数凹函数中最能提升稀疏性, 能消除局部极值点^[23]。因此本文采用拉普拉斯分布来对信号 $\mathbf{x}$ 进行建模, 对参数采用 Gamma 分布进行建模。设 β 的逆服从参数为 (a, b) 的 Gamma 分布:

$$\begin{aligned} p(\beta^{-1}) &= \text{Gamma}(\beta^{-1} | a, b) \\ &= \frac{b^a}{G(a)} (\beta^{-1})^{a-1} \exp(-b\beta^{-1}), \quad a, b \geq 0 \end{aligned} \quad (2)$$

其中 $G(a) = \int_0^\infty t^{a-1} e^{-t} dt$ 。又因为拉普拉斯分布与高斯似然函数混合后不可积。类似文献[12]采用对拉普拉斯分布的分层设计来克服此问题, 设中间变量为 γ_i , 其先验分布设为

$$\begin{aligned} p(\gamma_i | \lambda) &= \text{Gamma}(\gamma_i | 1, \lambda/2) \\ &= \frac{\lambda}{2} \exp \left\{ -\frac{\lambda}{2} \gamma_i \right\}, \quad \gamma_i \geq 0, \lambda \geq 0 \end{aligned} \quad (3)$$

给定 $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_N]^T$ 的条件下可得先验分布:

$$\begin{aligned} p(\mathbf{x} | \lambda) &= \int p(\mathbf{x} | \mathbf{g}) p(\mathbf{g} | \lambda) d\mathbf{g} \\ &= \left(\frac{\sqrt{\lambda}}{2} \right)^N \exp \left\{ -\sqrt{\lambda} \sum_{i=1}^N |x_i| \right\} \end{aligned} \quad (4)$$

为估计出 λ , 设其服从 Gamma 分布:

$$p(\lambda) = \text{Gamma} \left(\lambda \mid \frac{\xi}{2}, \frac{\xi}{2} \right) \quad (5)$$

结合上述分层贝叶斯模型可知数据、信号以及各参数的联合概率密度函数可表示为

$$\begin{aligned} p(\mathbf{y}, \mathbf{x}, \beta, \boldsymbol{\gamma}, \lambda) \\ = p(\mathbf{y} | \mathbf{x}, \beta) p(\mathbf{x} | \boldsymbol{\gamma}) p(\boldsymbol{\gamma} | \lambda) p(\lambda) p(\beta) \end{aligned} \quad (6)$$

2.2 贝叶斯推理

联合概率密度函数式(6)得出的信号后验概率分布是不可解析表达的, 需采用一些近似的办法来进行贝叶斯推理。本文采用第 2 类最大似然法。第 2 类似然函数由联合概率密度函数对信号的边缘化积分得到:

$$\begin{aligned} p(\mathbf{y}, \beta, \boldsymbol{\gamma}, \lambda) \\ &= \int p(\mathbf{y}, \mathbf{x}, \beta, \boldsymbol{\gamma}, \lambda) d\mathbf{x} \\ &= \int p(\mathbf{y} | \mathbf{x}, \beta) p(\mathbf{x} | \boldsymbol{\gamma}) p(\boldsymbol{\gamma} | \lambda) p(\lambda) p(\beta) d\mathbf{x} \\ &= (2\pi)^{-M/2} |\boldsymbol{\Sigma}_{\mathbf{y}}|^{-1/2} \cdot \exp \left\{ -\frac{1}{2} \mathbf{y}^T \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y} \right\} \\ &\quad \cdot p(\boldsymbol{\gamma} | \lambda) p(\lambda) p(\beta) \end{aligned} \quad (7)$$

其中 $\Sigma_y = \beta \mathbf{I}_M + \Phi \Gamma \Phi^T$, $\Gamma = \text{diag}\{\gamma\}$ 。本文通过对此似然函数的最大化, 将获得对参数 β, γ, λ 的估计。而由于信号 $\mathbf{x}$ 的后验概率分布为 $p(\mathbf{x} | \beta, \gamma, \lambda, \mathbf{y}) \propto p(\mathbf{x}, \beta, \gamma, \lambda, \mathbf{y})$, 当参数和观测数据给定时, 其将变换为多变量高斯分布 $\mathcal{N}(\mathbf{x} | \mu, \Sigma)$:

$$\mu \triangleq \mathbb{E}[\mathbf{x}; \beta, \gamma] = \Gamma \Phi^T \Sigma_y^{-1} \mathbf{y} \quad (8)$$

$$\Sigma \triangleq \text{Cov}[\mathbf{x}; \beta, \gamma] = \Gamma - \Gamma \Phi^T \Sigma_y^{-1} \Phi \Gamma \quad (9)$$

因此, 估计出最优的参数 β, γ, λ 后代入式(8)即可得到信号 $\mathbf{x}$ 的估计。最终问题等效为最大化式(7)时的参数最佳估计问题。当噪声功率与参数均设为均匀分布时, 该模型退化为 SBL 模型。

2.3 等效优化目标函数

为计算方便, 取式(7)的对数作为优化的目标函数, 去掉无关的常数项得

$$\begin{aligned} \mathcal{L}'(\beta, \gamma, \lambda, \xi) = & -\frac{1}{2} \lg |\Sigma_y| - \frac{1}{2} \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} + N \lg \frac{\lambda}{2} \\ & - \frac{\lambda}{2} \sum_i \gamma_i + (a-1) \lg \beta - b\beta \\ & + \frac{\xi}{2} \lg \frac{\xi}{2} - \lg G\left(\frac{\xi}{2}\right) \\ & + \left(\frac{\xi}{2} - 1\right) \lg \lambda - \frac{\xi}{2} \lambda \end{aligned} \quad (10)$$

根据 $\arg \max_{\beta, \gamma, \lambda, \xi} p(\mathbf{y}, \beta, \gamma, \lambda) = \arg \min_{\beta, \gamma, \lambda, \xi} -2\mathcal{L}'(\beta, \gamma, \lambda, \xi)$, 可得

$$\begin{aligned} \mathcal{L}(\beta, \gamma, \lambda, \xi) = & -2\mathcal{L}'(\beta, \gamma, \lambda, \xi) \\ = & \lg |\Sigma_y| + \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} - 2N \lg \frac{\lambda}{2} \\ & + \lambda \sum_i \gamma_i - 2(a-1) \lg \beta + 2b\beta \\ & - \xi \lg \frac{\xi}{2} + 2 \lg G\left(\frac{\xi}{2}\right) \\ & - (\xi - 2) \lg \lambda + \xi \lambda \end{aligned} \quad (11)$$

根据文献[12]的仿真分析可知, 在噪声功率假设为均匀分布时的重构性能要优于其它分布假设的性能; 若此时将参数 λ 也设为均匀分布, 则该目标函数将简化为 SBL 目标函数, 它拥有更简单的形式:

$$\mathcal{L}(\beta, \gamma) = \lg |\Sigma_y| + \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} \quad (12)$$

3 迭代重加权算法推导

由于目标函数式(12)是非凸多模的函数, 所以该问题是一个非凸优化问题。本文采用构造辅助函数进行优化的思想(MM 思想)。对凹函数进行凹共轭表达, 通过二次型的变分表示得到优化函数的一个上界函数, 该上界函数即为优化的辅助函数, 然后对此辅助函数进行交替的优化求解。

3.1 辅助函数

将 β, γ 视为 $N+1$ 维的整体参数, 由于 $\lg |\Sigma_y|$ 为参数 (β, γ^T) 的凹增函数, 由凸分析理论可知, 它可

用其共轭函数表示为

$$\lg |\Sigma_y| = \min_{z_0 \in \mathbb{R}_+^1, \mathbf{z} \in \mathbb{R}_+^N} (z_0 \beta + \mathbf{z}^T \gamma - g^*(z_0, \mathbf{z})) \quad (13)$$

其中 $g^*(z_0, \mathbf{z})$ 为 $\lg |\Sigma_y|$ 的凹共轭函数。

$$g^*(z_0, \mathbf{z}) = \min_{\beta \in \mathbb{R}_+^1, \gamma \in \mathbb{R}_+^N} (z_0 \beta + \mathbf{z}^T \gamma - \lg |\Sigma_y|) \quad (14)$$

由此可知

$$\begin{aligned} \mathcal{L}(\beta, \gamma) & \leq \mathcal{L}(\beta, \gamma, z_0, \mathbf{z}) \\ & = z_0 \beta + \mathbf{z}^T \gamma - g^*(z_0, \mathbf{z}) + \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} \end{aligned} \quad (15)$$

对于给定的参数 (β, γ^T) , 使式(15)等号成立的 $(z_0, \mathbf{z}^T)^T$ 为 $\nabla_{(\beta, \gamma)} \lg |\Sigma_y|$ 。根据求导法则有 $z_0 = \text{Tr}(\Sigma_y^{-1})$, $z_i = \text{Tr}(\Sigma_y^{-1} \phi_i \phi_i^T) = \phi_i^T \Sigma_y^{-1} \phi_i$, 也即 $\mathbf{z} = \text{diag}\{\Phi^T \Sigma_y^{-1} \Phi\}$ 。于是得到求解参数 β, γ 的迭代算法如表 1 所示。

表 1 求解参数 β, γ 的算法

局部算法: 固定参数 λ, ξ 求解 β, γ

步骤 1 初始化 $z_0 \in \mathbb{R}^1, \mathbf{z} \in \mathbb{R}^N$

步骤 2 对于给定的 $z_0, \mathbf{z}$, 对目标函数 $\mathcal{L}(\beta, \gamma)$ 的上界 $\mathcal{L}(\beta, \gamma, z_0, \mathbf{z})$ 进行优化, 求得最优的 β, γ :

$$(\hat{\beta}, \hat{\gamma}) = \arg \min_{\beta, \gamma} (z_0 \beta + \mathbf{z}^T \gamma + \mathbf{y}^T \Sigma_y^{-1} \mathbf{y}) \quad (16)$$

步骤 3 对于给定的 β, γ , 对目标函数 $\mathcal{L}(\beta, \gamma)$ 的上界 $\mathcal{L}(\beta, \gamma, z_0, \mathbf{z})$ 进行优化得到最紧的最优的 $z_0, \mathbf{z}$:

$$z_0^{\text{opt}} = \text{Tr}(\Sigma_y^{-1}), \mathbf{z}^{\text{opt}} = \nabla_{\gamma} \ln |\Sigma_y| = \text{diag}\{\Phi^T \Sigma_y^{-1} \Phi\} \quad (17)$$

循环运行步骤 2, 步骤 3 直至收敛。

3.2 参数的交替迭代更新

由于二次型 $\mathbf{y}^T \Sigma_y^{-1} \mathbf{y}$ 可变分表示为

$$\mathbf{y}^T \Sigma_y^{-1} \mathbf{y} = \min_{\mathbf{x} \in \mathbb{R}^N} (\beta^{-1} \|\mathbf{y} - \Phi \mathbf{x}\|^2 + \mathbf{x}^T \Gamma^{-1} \mathbf{x}) \quad (18)$$

其中 $\mathbf{x} = \mu = \Gamma \Phi^T \Sigma_y^{-1} \mathbf{y}$ 时, $\beta^{-1} \|\mathbf{y} - \Phi \mathbf{x}\|^2 + \mathbf{x}^T \Gamma^{-1} \mathbf{x}$ 取得极小值。代入式(16)可得

$$\begin{aligned} (\hat{\beta}, \hat{\gamma}) = & \arg \min_{\beta, \gamma} \min_{\mathbf{x}} (z_0 \beta + \mathbf{z}^T \gamma \\ & + \beta^{-1} \|\mathbf{y} - \Phi \mathbf{x}\|^2 + \mathbf{x}^T \Gamma^{-1} \mathbf{x}) \\ = & \arg \min_{\beta, \gamma} \min_{\mathbf{x}} \mathcal{L}(\beta, \gamma, z_0, \mathbf{z}, \mathbf{x}) \end{aligned} \quad (19)$$

而上界函数为 (β, γ) 和 $\mathbf{x}$ 的联合凸函数, 故可交换优化次序, 先优化 (β, γ) , 通过分别对目标函数求导置 0 可得

$$\beta^{\text{opt}} = \frac{\|\mathbf{y} - \Phi \mathbf{x}\|}{\sqrt{z_0}}, \quad \gamma_i^{\text{opt}} = \frac{|x_i|}{\sqrt{z_i}} \quad (20)$$

代入式(19)可得等效的优化问题为

$$\begin{aligned} \hat{\mathbf{x}} = & \arg \min_{\mathbf{x}} \left(2\sqrt{z_0} \|\mathbf{y} - \Phi \mathbf{x}\| + 2 \sum_i \sqrt{z_i} |x_i| \right) \\ = & \arg \min_{\mathbf{x}} \left(\sqrt{z_0} \|\mathbf{y} - \Phi \mathbf{x}\| + \sum_i \sqrt{z_i} |x_i| \right) \end{aligned} \quad (21)$$

此凸优化问题可通过凸优化工具箱即进行求解；再根据式(20)求得最优参数 (β, γ) 。则算法完整流程如表 2 所示。

表 2 参数自动调整迭代重加权算法

参数自动调整迭代重加权算法：求解稀疏贝叶斯重构问题

步骤 1 初始化 $z_0 \in \mathbb{R}_+^1, z \in \mathbb{R}_+^N$

步骤 2 对于给定的 z_0, z ，求解加权 L1 范数最小化问题

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left(\sqrt{z_0} \|\mathbf{y} - \Phi \mathbf{x}\| + \sum_i \sqrt{z_i} |x_i| \right) \quad (22)$$

步骤 3 根据式(20)对 (β, γ) 进行更新；

步骤 4 对于给定的参数 (β, γ) ，根据式(17)更新 z_0, z ；

循环运行步骤 2~步骤 4 直至收敛。

4 算法收敛性分析

根据 MM 算法的收敛性或全局收敛定理均可证明该算法收敛到局部最小点(或鞍点)。对噪声方差在收敛时的估计性能分析：用到对观测数据 $\mathbf{y}$ 的协方差矩阵 Σ_y 的分析，这里利用紧凑奇异值分解 $\Phi \hat{\Gamma}^{1/2} = \mathbf{U}_1 \Lambda \mathbf{V}_1^T$ 。

$$\begin{aligned} \Sigma_y &= \beta \mathbf{I}_M + \Phi \hat{\Gamma} \Phi^T \\ &= \hat{\beta} \mathbf{I}_M + \mathbf{U}_1 \Lambda \mathbf{V}_1^T \mathbf{V}_1 \Lambda \mathbf{U}_1^T \\ &= \hat{\beta} \mathbf{I}_M + \mathbf{U}_1 \Lambda^2 \mathbf{U}_1^T \\ &= [\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \hat{\beta} \mathbf{I}_K + \Lambda^2 & 0 \\ 0 & \hat{\beta} \mathbf{I}_{M-K} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{bmatrix} \end{aligned} \quad (23)$$

从而可得

$$\Sigma_y^{-1} = [\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} (\hat{\beta} \mathbf{I}_K + \Lambda^2)^{-1} & 0 \\ 0 & 1/\hat{\beta} \mathbf{I}_{M-K} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{bmatrix} \quad (24)$$

亦有

$$\text{Tr}(\Sigma_y)^{-1} = \sum_{i=1}^K \frac{1}{\hat{\beta} + \Lambda_i^2} + \sum_{i=K+1}^M \frac{1}{\hat{\beta}} \quad (25)$$

设在 t 步迭代时各参数表示为 $\mathbf{x}^{(t)}$ ，

$$\begin{aligned} \beta^* &= \frac{\|\mathbf{y} - \Phi \mathbf{x}^*\|}{\sqrt{z_0^*}} = \frac{\|\mathbf{y} - \Phi \mathbf{x}^*\|}{\sqrt{\text{Tr}(\Sigma_y^*)^{-1}}} \\ &= \frac{\|\mathbf{y} - \Phi \mathbf{x}^*\|}{\sqrt{\sum_{i=1}^K \frac{1}{\beta^* + \Lambda_i^2} + \sum_{i=K+1}^M \frac{1}{\beta^*}}} \\ &= \frac{\|\mathbf{y} - \Phi \mathbf{x}^*\|^2}{\sum_{i=1}^K \frac{1}{1 + \Lambda_i^2/\beta^*} + M - K} \end{aligned} \quad (26)$$

由此可得收敛时噪声方差估计的上下界：

$$\begin{aligned} \frac{\|\mathbf{y} - \Phi \mathbf{x}^*\|^2}{M} &\leq \beta^* = \frac{\|\mathbf{y} - \Phi \mathbf{x}^*\|^2}{\sum_{i=1}^K \frac{1}{1 + \Lambda_i^2/\beta^*} + M - K} \\ &\leq \frac{\|\mathbf{y} - \Phi \mathbf{x}^*\|^2}{M - K} \end{aligned} \quad (27)$$

即在高信噪比以及稀疏度估计准确时对噪声功率的估计可接近理想估计：

$$\beta^{\text{oracal}} = \frac{\|\mathbf{y} - \Phi \mathbf{x}^{\text{oracal}}\|^2}{M - K} \quad (28)$$

5 实验及分析

在稀疏求解算法中，Wipf 等的迭代 L1, L2 算法由于采用迭代重加权的策略，其重构性能要较其它算法要好，但是该算法仍然需要预先获知噪声的方差^[18]。本文选择与 Wipf 等的迭代 L1, L2 算法进行比较。

5.1 实验 1

实验数据产生方式：稀疏向量 $\mathbf{x} \in \mathbb{R}^{256 \times 1}$ 维数为 $N = 256$ ，稀疏度为 $K < 128$ ，给定稀疏度为 K 后，长为 K 的支撑集在 $\{1, 2, \dots, N\}$ 上均匀随机地产生，并在此支撑集上使信号的幅度为 1；观测矩阵 Φ 为 100×256 ，其中元素由服从独立同分布的标准高斯变量的一次实现，再对矩阵的每个列向量进行单位化产生。噪声为高斯白噪声，其方差为 0.01。观测向量 $\mathbf{y} \in \mathbb{R}^{100 \times 1}$ 维数为 100，由常用模型 $\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}$ 产生。实验安排如表 3 所示。

表 3 实验目的与参数设置说明

对比算法	比较不同算法的重构性能与稀疏度的关系		比较不同算法的重构性能与迭代步数的关系	
	实验 1.1	实验 1.2	实验 1.3	实验 1.4
本文算法	不预设参数	不预设参数	不预设参数	不预设参数
Wipf 等的迭代 L1 算法	非理想参数设置	理想参数设置	非理想参数设置	理想参数设置
Wipf 等的迭代 L2 算法	非理想参数设置	理想参数设置	非理想参数设置	理想参数设置

实验 1.1 稀疏度 $K=1:5:60$, 每个稀疏度重复 100 次, 误差用 L2 范数相对比值给出, 并计算获得平均值。本文算法自适应调整参数, Wipf 迭代 L1, L2 算法均设定 $\lambda = 0.1$ 。实验结果如图 1 所示。

实验 1.2 实验条件同实验 1.1, 不同的是 Wipf 迭代 L1, L2 算法均为理想参数设置。实验结果如图 2 所示。

实验 1.3 稀疏度 $K=5:5:30$, 每个稀疏度、每个迭代步数重复 100 次, 误差用 L2 范数相对比值给出, 并计算获得平均值。本文算法自适应调整参数, Wipf 迭代 L1, L2 算法均设定 $\lambda = 0.1$ 。实验结果如图 3 所示(由于曲线太多, 无法区分, 仅保留稀疏度为 $K=5$ 和 $K=30$ 的结果进行对比)。

实验 1.4 实验条件同实验 1.3, 不同的是 Wipf 迭代 L1, L2 算法均为理想参数设置。实验结果如图 4 所示。

由图可知, 在 Wipf 等迭代重加权 L1, L2 范数最小化选择非理想参数时, 本文算法的重构性能要明显优于这两种算法。如图 1 所示, 稀疏度小于 25 时, 本文算法的重构误差要优于 Wipf 的两种算法近 100 倍, 稀疏度增大时, 其差距逐渐减小; 而在重构误差与迭代次数关系实验中, 如图 3 所示, 本文

算法在第 2 步迭代就达到了比非理想参数设置的 Wipf 方法更好的重构结果; Wipf 等算法在选择理想参数时, 如图 2, 图 4 所示, 本文算法与后两种算法的性能相当, 甚至更优。实验结果显示了参数自动调整算法的明显优势。

5.2 实验 2

实验 2.1 为更加直观地对比算法性能, 运用本文算法和 Wipf 迭代 L1 算法对一个稀疏的 1 维信号进行重构。实验数据产生方式同实验 1.1, Wipf 迭代 L1 算法为非理想参数设置, 本文方法不预设参数。不同的是稀疏度 $K=5$, 取值服从均值为 0, 方差为 1 的高斯分布。迭代步数为 10 步, 重构结果如图 5 所示。

如图 5(a), 5(b)所示, 在迭代步数定为 10 步时, 两算法效果差距明显。Wipf 迭代 L1 算法结果中, 不仅非零值的位置与原信号有所不同, 取值也有了较大的偏差, 而本文算法能够很好地恢复出原信号。

实验 2.2 为检验算法效率, 在实验 1.1 的基础上分别对本文算法, Wipf 迭代 L1, L2 算法进行时间对比, 其中迭代步数为 15, 稀疏度 $K=60$ 。结果如表 4 所示。

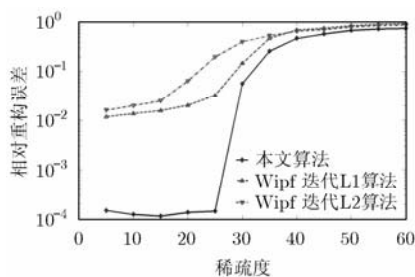


图1 重构误差与稀疏度的关系比较
(Wipf 非理想参数设置)

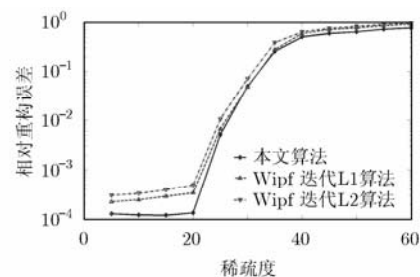


图2 重构误差与稀疏度的关系比较
(Wipf 理想参数设置)

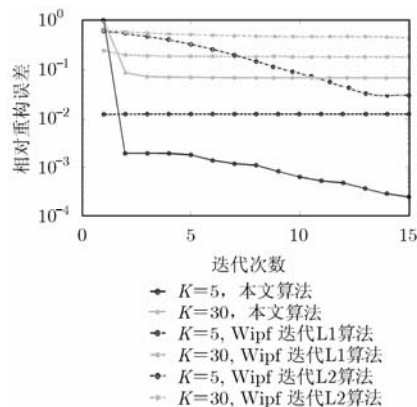


图3 重构误差与迭代次数的关系比较
(Wipf 非理想参数设置)

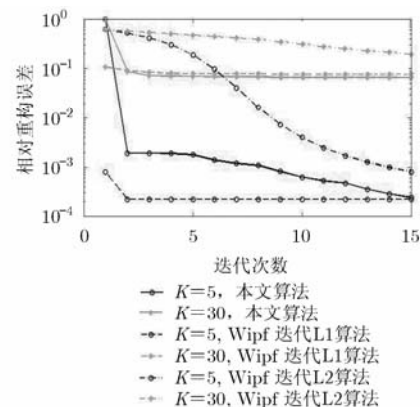


图4 重构误差与迭代次数的关系比较
(Wipf 理想参数设置)

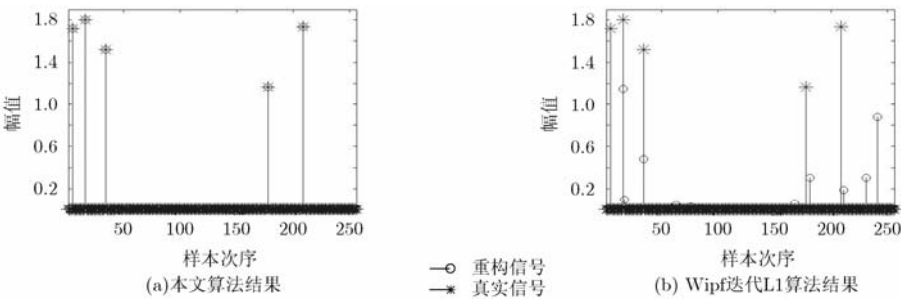


图 5 信号重构结果对比

表 4 处理时间对比(s)

数据集	Wipf 等的迭 代 L1 算法	Wipf 等的迭 代 L2 算法	本文算法
仿真数据集	25.46	1.03	20.71

从表 4 看出，由于 Wipf 迭代 L2 算法在每一次迭代中都有解析解，综合效率最高，而 Wipf 迭代 L1 算法在每次迭代的目标函数中都增加了 L1 范数正则项，运算开销较大，而本文算法由于运用了贝叶斯学习的思想，运算时间也较长，但仍然较 Wipf 迭代 L1 算法效率要高。

6 结论

当缺乏系统噪声分布等先验信息时，现有的稀疏重构算法无法合理设置正则化参数，进而获得良好的重构性能。为解决这个问题，本文提出一种参数自动调整的迭代重加权混合 L2-L1 范数最小化算法。由理论推导和仿真实验可知，在理想参数已知时，该算法与 Wipf 迭代重加权 L1 及 L2 范数最小化算法的重构性能相当，甚至更优；而在选择非理想参数时，本文算法的重构性能要明显优于后两者算法。此外，本文算法在运算效率上也与其算法相当。算法收敛性分析表明，在收敛点，对噪声方差的估计位于对噪声方差的 oracal 估计和矩估计之间，在高信噪比时将接近 oracal 估计，这在统计意义上说明了本文算法的有效性。

参 考 文 献

[1] Candès E J, Romberg J, and Tao T. Stable signal recovery from incomplete and inaccurate measurements[J]. *Communications on Pure and Applied Mathematics*, 2006, 59(8): 1207-1233.

[2] Candès E J and Tao T. The Dantzig selector: statistical estimation when p is much larger than n [J]. *Annals of Statistics*, 2007, 35(6): 2358-2364.

[3] Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective[J]. *Journal of Royal Statistical Society Series*

B, 2011, 73(3): 273-282.

[4] Tropp J A and Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit[J]. *IEEE Transactions on Information Theory*, 2007, 53(12): 4655-4666.

[5] Donoho D L, Tsaig Y, Drori I, et al. Sparse solution of underdetermined systems of linear equations by Stagewise Orthogonal Matching Pursuit[J]. *IEEE Transactions on Information Theory*, 2012, 58(2): 1094-1121.

[6] Dai W and Milenkovic O. Subspace pursuit for compressive sensing signal reconstruction[J]. *IEEE Transactions on Information Theory*, 2009, 55(5): 2230-2249.

[7] Wipf D P. Bayesian methods for finding sparse representations[D]. [Ph. D. dissertation], UC, San Diego, 2006.

[8] Zhang Z and Rao B D. Recovery of block sparse signals using the framework of block sparse Bayesian learning[C]. *The 37th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, La Jolla, CA, USA, 2012: 3345-3348.

[9] 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.

[10] Qiu K and Dogandzic A. Variance-component based sparse signal reconstruction and model selection[J]. *IEEE Transactions on Signal Processing*, 2010, 58(6): 2935-2951.

[11] Zhang Z and Rao B D. Iterative reweighted algorithms for sparse signal recovery with temporally correlated source vectors[C]. *The 36th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, 2011: 3932-3935.

[12] Lin Qin, Guo Wei, Fu XueYang, et al. MR image reconstruction by path-based sparse representation[J]. *Journal of Theoretical and Applied Information Technology*, 2013, 49(1): 107-112.

[13] Wipf D and 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.
- [14] Zhang Z, Wang S, Liu D, *et al.* EP-GIG Priors and applications in Bayesian sparse learning[J]. *The Journal of Machine Learning Research*, 2012, 13(Jun): 2031–2061.
- [15] Carlin M. Directions-of-arrival estimation through Bayesian compressive sensing strategies[J]. *IEEE Transactions on Antennas and Propagation*, 2013, 61(7): 3828–3838.
- [16] Arberet S, Vanderghelynst P, Carrillo R, *et al.* Sparse Reverberant audio source separation via reweighted analysis [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2013, 21(7): 1391–1402.
- [17] Peng J, Zhu J, Han W, *et al.* Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer[J]. *The Annals of Applied Statistics*, 2010, 4(1): 53–77.
- [18] Zhao P, Rocha G, and Yu B. The composite absolute penalties family for grouped and hierarchical variable selection[J]. *The Annals of Statistics*, 2009, 37(6A): 3468–3497.
- [19] Jacob L, Obozinski G, and Vert J. Group lasso with overlap and graph lasso[C]. The 26th International Conference on Machine Learning, Montreal, Canada, 2009: 433–440.
- [20] Jenatton R, Mairal J, Obozinski G, *et al.* Proximal methods for sparse hierarchical dictionary learning[C]. The 27th International Conference on Machine Learning, Haifa, Israel, 2010: 487–494.
- [21] Liu J and Ye J. Moreau-Yosida regularization for grouped tree structure learning[C]. The 24th Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2010: 1459–1467.
- [22] Hao D N, Chuong L H, and Lesnic D. Heuristic regularization methods for numerical differentiation[J]. *Computers & Mathematics with Applications*, 2012, 63(4): 816–826.
- [23] Seeger M and Nickisch H. Compressed sensing and Bayesian experimental design[C]. International Conference on Machine Learning, Omni Press, 2008: 912–919.
- 夏建明: 男, 1982年生, 博士, 研究方向为模式识别、机器学习、压缩感知。
- 杨俊安: 男, 1965年生, 教授, 博士生导师, 研究方向为信号处理、智能计算等。
- 陈 功: 男, 1982年生, 博士生, 研究方向为压缩感知、统计学习。