

利用期望-最大化算法实现基于动态词典的压缩感知

胡磊* 周剑雄 石志广 付强
(国防科技大学 ATR 实验室 长沙 410073)

摘要: 在现有压缩感知(CS)理论中, 重构信号需要预设其稀疏表示词典。对于以参数化模型表示的信号, 只能预知该词典为某种形式的参数化词典, 参数的具体取值难以确定。若将参数设定为取值空间的均匀离散格点, 预设词典与真实词典之间的失配将使传统 CS 重构方法的性能严重恶化。为解决这一问题, 该文提出一种基于动态词典的 CS 重构方法。通过迭代地优化词典参数, 该方法在信号重构过程中对词典进行动态调整。为同时实现稀疏恢复与词典调整, 该方法利用变分期望-最大化(EM)算法交替执行信号系数估计与词典参数优化。实验结果表明所提方法是有效的。

关键词: 压缩感知(CS); 稀疏恢复; 动态词典; 期望-最大化(EM); 变分贝叶斯近似

中图分类号: TN911.7

文献标识码: A

文章编号: 1009-5896(2012)11-2554-07

DOI: 10.3724/SP.J.1146.2012.00347

An EM-based Approach for Compressed Sensing Using Dynamic Dictionaries

Hu Lei Zhou Jian-xiong Shi Zhi-guang Fu Qiang

(ATR Lab, National University of Defense Technology, Changsha 410073, China)

Abstract: In the current Compressed Sensing (CS) theory, signal reconstruction depends on presetting an appropriate sparsifying dictionary. For signals characterized by parametric models, this dictionary is known to be a parameterized dictionary of a certain form, but the values of the parameters are difficult to determine. If the parameters are set to a group of uniform grid points, the mismatch between the assumed and the actual sparsifying dictionaries will cause the performance of conventional CS reconstruction methods to degrade considerably. To address this, a CS reconstruction method that utilizes dynamic dictionaries is proposed. By iteratively optimizing dictionary parameters, the method refines the dictionary dynamically during signal reconstruction. To achieve joint sparse recovery and dictionary refinement, the method alternates between steps of signal coefficients estimation and dictionary parameters optimization under the framework of the variational Expectation-Maximization (EM) algorithm. Experimental results demonstrate the effectiveness of the proposed method.

Key words: Compressed Sensing (CS); Sparse recovery; Dynamic dictionaries; Expectation-Maximization (EM); Variational Bayesian approximation

1 引言

压缩感知(Compressed Sensing, CS)是一种近年兴起的基础信号处理理论, 其基本结论是: 稀疏信号可从其高度欠采样样本中以高精度恢复^[1-3]。基于这一理论, 信号处理系统有望突破传统奈奎斯特采样率的限制, 这对降低系统采样率、数据量与存储量具有重要意义。

在现有 CS 理论中, 实现信号重构需要预先设定一个合适的稀疏表示词典。然而, 对于以参数化模型表示的信号, 只能预知其在某种形式的参数化

词典上稀疏, 而词典参数的具体取值则难以确定。目前, 为实现这类信号的重构, 一般将词典参数设定为参数空间的某组均匀离散格点。事实上, 只有当原始信号的真实参数恰好位于预设的参数格点上时, 该信号才在所设词典上稀疏; 否则, 当信号参数偏离预设格点时, 利用所设词典表示信号将存在能量的格点间泄漏, 从而导致信号在所设词典上并非真正稀疏。对实际信号, 由于其所含参数事先未知, 不可能预设一组格点以保证其参数恰好位于格点上, 因此通过盲目离散参数空间形成的词典与信号的真实稀疏表示词典之间必然存在失配。已有研究表明, 当上述词典失配存在时, 传统 CS 重构方法的性能将严重恶化^[4]。为解决这一问题, 本文提出一种基于动态词典的 CS 信号重构方法。通过迭

2012-03-29 收到, 2012-09-03 改回

国家自然科学基金(60972113, 61101179)资助课题

*通信作者: 胡磊 vanilla_hugh_hl@163.com

代地优化词典参数，该方法将词典的动态调整融入到信号重构过程中。为使稀疏恢复与词典优化同步进行，该方法利用期望-最大化 (Expectation-Maximization, EM) 算法迭代地执行以下两个步骤：(1) 给定词典下的信号稀疏系数恢复；(2) 给定稀疏系数下的信号词典更新。为验证所提方法，将其应用于谱压缩感知领域，实验结果表明该方法是有用的。

2 信号与测量模型

设原始信号 x_n 可用如下所示的参数化模型描述：

$$x_n = \sum_{k=1}^K a_k s(t_n, \theta_k), \quad n = 1, \dots, N \quad (1)$$

式中， K 表示模型阶数， $s(t, \theta)$ 表示以 θ 为参数的复数域波形， $a_k \in \mathbb{C}$ 与 $\theta_k \in \mathbb{R}^d$ 分别表示第 k 个分量信号的复幅度与参数取值。可以看出，上述模型涵盖了工程领域需要处理的多种信号形式，如谐波信号、多分量线性调频信号以及多项式相位信号等。对任意参数取值 $\theta \in \mathbb{R}^d$ ，令：

$$e(\theta) = [s(t_1, \theta), \dots, s(t_N, \theta)]^T \quad (2)$$

则信号 x_n ($n = 1, \dots, N$) 可表示为如下向量形式：

$$\mathbf{x} = \mathbf{F}(\theta_1, \dots, \theta_K) \cdot \mathbf{a} \quad (3)$$

式中 $\mathbf{x} = [x_1, \dots, x_N]^T$ ， $\mathbf{a} = [a_1, \dots, a_K]^T$ ，模型矩阵 $\mathbf{F}(\theta_1, \dots, \theta_K)$ 表示为

$$\mathbf{F}(\theta_1, \dots, \theta_K) = [\mathbf{e}(\theta_1) \ \dots \ \mathbf{e}(\theta_K)] \quad (4)$$

在 CS 理论中，信号测量通过如下模型获得

$$\mathbf{y} = \Phi \mathbf{x} + \varepsilon \quad (5)$$

式中 $\Phi \in \mathbb{C}^{M \times N}$ 表示测量矩阵，且 $M \ll N$ ， $\varepsilon \in \mathbb{C}^M$ 表示测量噪声， $\mathbf{y} \in \mathbb{C}^M$ 表示 \mathbf{x} 的压缩测量。在 CS 理论中，信号重构是通过稀疏恢复算法实现的。基于此，欲利用 \mathbf{y} 重构出 \mathbf{x} ，需要设定 \mathbf{x} 的稀疏表示词典。在现有研究中，一般将这一词典设定为参数空间中一组均匀离散格点所对应的波形匹配词典。对任意一组参数离散格点 $\tilde{\theta} = \{\tilde{\theta}_1, \dots, \tilde{\theta}_P\}$ ，存在一个相应的波形匹配词典 $\mathbf{D}(\tilde{\theta})$ ，定义为

$$\mathbf{D}(\tilde{\theta}) = [\mathbf{e}(\tilde{\theta}_1) \ \dots \ \mathbf{e}(\tilde{\theta}_P)] \quad (6)$$

由式(1)可以看出，欲使 \mathbf{x} 在 $\mathbf{D}(\tilde{\theta})$ 上真正稀疏，必须保证其分量参数 $\{\theta_1, \dots, \theta_K\}$ 全部位于格点 $\tilde{\theta}$ 上，显然这一条件对实际信号无法满足。为解决这一问题，本文考虑在稀疏恢复过程中采用动态词典，基本思路是：将 $\tilde{\theta}$ 看作 $\mathbf{D}(\tilde{\theta})$ 的可变参数，从而可通过不断调整 $\tilde{\theta}$ 来实现 $\mathbf{D}(\tilde{\theta})$ 的自适应优化。

3 基于动态词典的信号重构

设 \mathbf{x} 在 $\mathbf{D}(\tilde{\theta})$ 上的稀疏表示系数为 \mathbf{w} ，则测量模

型式(5)可表示为

$$\mathbf{y} = \Phi \mathbf{D}(\tilde{\theta}) \mathbf{w} + \varepsilon \quad (7)$$

欲重构 \mathbf{x} ，需要同时进行系数 \mathbf{w} 的恢复与词典 $\mathbf{D}(\tilde{\theta})$ 的优化。为解决这一问题，可迭代执行下述两个步骤：(1) 给定词典，恢复稀疏表示系数；(2) 固定稀疏表示系数，更新词典。由于词典更新是通过参数 $\tilde{\theta}$ 的调整实现的，因此上述流程事实上是一个交替优化 \mathbf{w} 与 $\tilde{\theta}$ 的过程。可以看出，这一过程类似于 EM 算法^[5]的参数估计流程，因此可在这一算法框架下实现上述步骤。

3.1 变分 EM 算法

为求解未知参数的最大似然估计，EM 算法迭代地执行两个步骤，即期望步 (Expectation Step) 与最大化步 (Maximization Step)。在期望步，算法将观测模型中的隐含未知量视作随机变量，基于观测与当前参数估计求解其数学期望；在最大化步，算法基于期望步所得隐含变量期望更新未知参数的最大似然估计^[5]。然而，对许多复杂模型，由于隐含变量的后验概率分布无法解析求解，传统 EM 算法不再适用。为解决这一问题，统计推理领域新近兴起的变分近似 (variational approximation) 技术^[6]被应用于 EM 算法的原理与实现方式推导之中，由此衍生出一种基于变分法的新型 EM 算法，称为变分 EM^[7]。假设所考虑模型的观测为 \mathbf{y} ，待估计参数为 $\boldsymbol{\eta}$ ，隐含变量为 \mathbf{z} ，以 $p(\mathbf{y}, \mathbf{z}; \boldsymbol{\eta})$ 表示全数据 (观测 + 隐含变量) 的似然函数。在期望步，变分 EM 算法基于 \mathbf{y} 与 $\boldsymbol{\eta}$ 的当前估计 $\boldsymbol{\eta}^{\text{OLD}}$ 计算 \mathbf{z} 的后验期望，为此通过式(8)得到 \mathbf{z} 的各个分量的后验概率分布^[7]：

$$\ln q_k = \left\langle \ln p(\mathbf{y}, \mathbf{z}; \boldsymbol{\eta}^{\text{OLD}}) \right\rangle_{i=k} + C, \quad (8)$$

式(8)中 q_k 表示 \mathbf{z} 的第 k 个分量 z_k 的后验概率密度， $\langle \cdot \rangle_{i=k}$ 表示相对于 q_i ($i \neq k$) 的数学期望， C 表示归一化常数。在最大化步，算法通过求解式(9)更新 $\boldsymbol{\eta}$ 的估计^[7]：

$$\boldsymbol{\eta}^{\text{NEW}} = \arg \max_{\boldsymbol{\eta}} \langle \ln p(\mathbf{y}, \mathbf{z}; \boldsymbol{\eta}) \rangle_{q(\mathbf{z})} \quad (9)$$

式中 $q(\mathbf{z}) = \prod_k q_k$ ， $\langle \cdot \rangle_{q(\mathbf{z})}$ 表示相对于 $q(\mathbf{z})$ 的数学期望。

3.2 稀疏表示系数的更新方法^[7]

将变分 EM 算法应用于式(7)对应的模型，可将 \mathbf{w} 建模为随机变量，将 $\tilde{\theta}$ 视为确定性未知参数，从而可利用变分 EM 算法的期望步与最大化步对两者进行交替更新。为此，需要首先为 \mathbf{w} 设定合适的先验分布。考虑到 \mathbf{w} 具有稀疏特性，假设其服从如下形式的分层复高斯先验^[7]：

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \frac{1}{\pi^P |\mathbf{A}|^{-1}} \exp(-\mathbf{w}^H \mathbf{A} \mathbf{w}) \quad (10)$$

$$p(\boldsymbol{\alpha}; a, b) = \prod_{i=1}^P \text{Gamma}(\alpha_i | a, b) \quad (11)$$

式中 α_i^{-1} 表示 \boldsymbol{w} 的第 i 个分量的先验方差, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_P]^T$, $\mathbf{A} = \text{diag}(\boldsymbol{\alpha})$, $\text{Gamma}(\alpha_i | a, b)$ 表示以 a, b 为参数的 Gamma 分布。事实上, 分层高斯 (hierarchical Gauss) 分布是稀疏贝叶斯学习 (Sparse Bayesian Learning, SBL) 领域所采用的典型稀疏系数先验分布^[8], 它能够诱导稀疏性, 并且是高斯概率密度的共轭分布。假设测量噪声为独立同分布的零均值复高斯噪声, 设其方差为 β^{-1} , 则有

$$p(\boldsymbol{\varepsilon} | \beta) = (\pi\beta^{-1})^{-M} \exp(-\beta \|\boldsymbol{\varepsilon}\|_2^2) \quad (12)$$

将 β 建模为随机变量, 且假设其服从 Gamma 先验分布, 即

$$p(\beta; c, d) = \text{Gamma}(\beta | c, d) \quad (13)$$

设 $\tilde{\boldsymbol{\theta}}$ 的当前值为 $\tilde{\boldsymbol{\theta}}^{\text{OLD}}$, 其对应的稀疏表示词典为 $\mathbf{D}(\tilde{\boldsymbol{\theta}}^{\text{OLD}})$, 根据式(7)与式(12), 期望步所对应的似然函数为

$$\begin{aligned} p(\mathbf{y} | \boldsymbol{w}, \beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) \\ = (\pi\beta^{-1})^{-M} \exp\left(-\beta \left\| \mathbf{y} - \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}^{\text{OLD}}) \boldsymbol{w} \right\|_2^2\right) \end{aligned} \quad (14)$$

根据式(8), 隐含变量 \boldsymbol{w} , $\boldsymbol{\alpha}$ 及 β 的后验概率分布通过下式得到

$$\begin{aligned} \ln q(\boldsymbol{w}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) \\ = \left\langle \ln p(\mathbf{y} | \boldsymbol{w}, \beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) p(\boldsymbol{w} | \boldsymbol{\alpha}) \right\rangle_{q(\boldsymbol{\alpha})q(\beta)} + C \end{aligned} \quad (15)$$

$$\ln q(\boldsymbol{\alpha}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) = \left\langle \ln p(\boldsymbol{w} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha}) \right\rangle_{q(\boldsymbol{w})q(\beta)} + C \quad (16)$$

$$\begin{aligned} \ln q(\beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) \\ = \left\langle \ln p(\mathbf{y} | \boldsymbol{w}, \beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) p(\beta) \right\rangle_{q(\boldsymbol{w})q(\boldsymbol{\alpha})} + C \end{aligned} \quad (17)$$

将式(10)与式(14)代入式(15), 经过推导可得

$$q(\boldsymbol{w}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) = \mathcal{CN}(\boldsymbol{w} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (18)$$

$$\boldsymbol{\mu} = \langle \beta \rangle \boldsymbol{\Sigma} \mathbf{D}^H(\tilde{\boldsymbol{\theta}}^{\text{OLD}}) \Phi^H \mathbf{y} \quad (19)$$

$$\boldsymbol{\Sigma} = \left(\langle \beta \rangle \mathbf{D}^H(\tilde{\boldsymbol{\theta}}^{\text{OLD}}) \Phi^H \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}^{\text{OLD}}) + \langle \mathbf{A} \rangle \right)^{-1} \quad (20)$$

其中, $\mathcal{CN}(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ 表示均值为 $\boldsymbol{\mu}$, 协方差矩阵为 $\boldsymbol{\Sigma}$ 的复高斯分布, $\langle \mathbf{A} \rangle = \text{diag}(\langle \boldsymbol{\alpha} \rangle)$, $\langle \boldsymbol{\alpha} \rangle = [\langle \alpha_1 \rangle, \dots, \langle \alpha_P \rangle]^T$ 。将式(10)与式(11)代入式(16), 经推导得到

$$q(\boldsymbol{\alpha}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) = \prod_{i=1}^P \text{Gamma}(\alpha_i | \tilde{a}_i, \tilde{b}_i) \quad (21)$$

$$\tilde{a}_i = a + 1 \quad (22)$$

$$\tilde{b}_i = |\mu_i|^2 + \Sigma_{ii} + b \quad (23)$$

将式(13)与式(14)代入式(17), 可得

$$q(\beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) = \text{Gamma}(\beta | \tilde{c}, \tilde{d}) \quad (24)$$

$$\tilde{c} = c + M \quad (25)$$

$$\begin{aligned} \tilde{d} = & \left\| \mathbf{y} - \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}^{\text{OLD}}) \boldsymbol{\mu} \right\|_2^2 \\ & + \text{trace} \left(\mathbf{D}^H(\tilde{\boldsymbol{\theta}}^{\text{OLD}}) \Phi^H \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}^{\text{OLD}}) \boldsymbol{\Sigma} \right) + d \end{aligned} \quad (26)$$

其中 $\text{trace}(\cdot)$ 表示矩阵的求迹运算。根据 Gamma 分布的性质, 可通过式(27), 式(28)计算得到 α_i 与 β 的后验期望:

$$\langle \alpha_i \rangle = \tilde{a}_i / \tilde{b}_i, \quad i = 1, \dots, P \quad (27)$$

$$\langle \beta \rangle = \tilde{c} / \tilde{d} \quad (28)$$

可以看出, \boldsymbol{w} , $\boldsymbol{\alpha}$ 及 β 的后验分布互相依赖, 因此, 可首先基于式(27)与式(28)以及 \boldsymbol{w} 的当前分布更新 $\boldsymbol{\alpha}$ 与 β 的后验期望, 再利用这些结果以及式(19), 式(20)更新 \boldsymbol{w} 的后验均值与协方差矩阵。在上述过程中, 参数 a, b, c, d 一般设置为接近于 0 的常数 (如 $a = b = c = d = 10^{-6}$), 这相当于假设 $\boldsymbol{\alpha}$ 与 β 服从无信息先验分布^[8]。

3.3 词典参数的更新方法

根据式(9), 参数 $\tilde{\boldsymbol{\theta}}$ 的更新通过式(29)实现

$$\begin{aligned} \tilde{\boldsymbol{\theta}}^{\text{NEW}} \\ = \arg \max_{\tilde{\boldsymbol{\theta}}} \left\langle \ln p(\mathbf{y}, \boldsymbol{w}, \boldsymbol{\alpha}, \beta; \tilde{\boldsymbol{\theta}}) \right\rangle_{q(\boldsymbol{w}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) q(\boldsymbol{\alpha}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) q(\beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}})} \end{aligned} \quad (29)$$

通过忽略目标函数中与 $\tilde{\boldsymbol{\theta}}$ 无关的项, 式(29)可简化为

$$\tilde{\boldsymbol{\theta}}^{\text{NEW}} = \arg \max_{\tilde{\boldsymbol{\theta}}} \left\langle \ln p(\mathbf{y} | \boldsymbol{w}, \beta; \tilde{\boldsymbol{\theta}}) \right\rangle_{q(\boldsymbol{w}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) q(\boldsymbol{\alpha}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) q(\beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}})} \quad (30)$$

根据

$$p(\mathbf{y} | \boldsymbol{w}, \beta; \tilde{\boldsymbol{\theta}}) = (\pi\beta^{-1})^{-M} \exp\left(-\beta \left\| \mathbf{y} - \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}) \boldsymbol{w} \right\|_2^2\right) \quad (31)$$

可将式(30)转化为

$$\tilde{\boldsymbol{\theta}}^{\text{NEW}} = \arg \min_{\tilde{\boldsymbol{\theta}}} \left\langle \left\| \mathbf{y} - \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}) \boldsymbol{w} \right\|_2^2 \right\rangle_{q(\boldsymbol{w}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) q(\boldsymbol{\alpha}; \tilde{\boldsymbol{\theta}}^{\text{OLD}}) q(\beta; \tilde{\boldsymbol{\theta}}^{\text{OLD}})} \quad (32)$$

进一步推导可得

$$\begin{aligned} \tilde{\boldsymbol{\theta}}^{\text{NEW}} = \arg \min_{\tilde{\boldsymbol{\theta}}} \left\{ \left\| \mathbf{y} - \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}) \boldsymbol{\mu} \right\|_2^2 \right. \\ \left. + \text{trace} \left(\mathbf{D}^H(\tilde{\boldsymbol{\theta}}) \Phi^H \Phi \mathbf{D}(\tilde{\boldsymbol{\theta}}) \boldsymbol{\Sigma} \right) \right\} \end{aligned} \quad (33)$$

这就是说, 词典参数格点通过求解式(33)所示的优化问题获得更新。

3.4 信号重构流程

建立了稀疏系数 \boldsymbol{w} 与词典参数 $\tilde{\boldsymbol{\theta}}$ 的更新流程之后, 原始信号 \boldsymbol{x} 的重构可通过交替执行这些流程实

现。类似于稀疏贝叶斯学习^[8]，随着上述过程中迭代次数的增加，大部分 $\langle \alpha_i \rangle$ 趋于较大值，这将导致大部分 μ_i 和 Σ_{ii} 趋近于 0，从而实现表示系数的稀疏性。基于这一原理，可在上述迭代过程中通过对 $\langle \alpha \rangle$ 进行门限监测实现参数格点的实时修剪。设 α_{Th} 表示某一门限，以 $\tilde{\theta}, \tilde{\theta}$ 分别表示修剪前与修剪后格点，则对 $\tilde{\theta}$ 中的任意元素 $\tilde{\theta}_i$ ，当且仅当 $\langle \alpha_i \rangle \leq \alpha_{Th}$ 时， $\tilde{\theta}_i \in \tilde{\theta}$ 。

基于以上分析，利用动态词典重构 x 的流程可概括如下：

(1)初始化：设置初始参数格点为 $\tilde{\theta}^{init}$ ，由此形成初始词典 $D(\tilde{\theta}^{init})$ ，令 $a = b = c = d = 10^{-6}$ ，计算 α 与 β 的先验期望并将其作为后验期望，利用式(19)与式(20)初始化 w 的后验均值与协方差矩阵；

(2)更新稀疏表示系数：根据 3.2 节所给方法更新稀疏表示系数 w ；

(3)修剪参数格点：根据上面介绍方法修剪当前参数格点，同时对 w 的后验均值与协方差矩阵作相应修剪；

(4)更新参数格点：根据式(33)对参数格点进行更新，并利用新格点形成新的稀疏表示词典；

(5)检验终止条件：如果终止条件满足，则转到第(6)步；否则，执行第(2)–(4)步；

(6)获得重构结果：利用当前稀疏表示系数与参数格点合成重构信号 \hat{x} 。

由于式(33)是一个非凸优化问题，参数格点的初始化会影响信号重构结果。一种自然的选择是将 $\tilde{\theta}^{init}$ 设置为参数空间的均匀离散格点。另外，文献[9]提出了一种基于 Fisher 信息的初始参数格点设计方法，利用该方法可将 $\tilde{\theta}^{init}$ 设计为非均匀格点。

上述算法进行一次迭代的计算负担主要决定于两个方面：一是系数 w 的后验协方差矩阵 Σ 的计算，二是新参数格点 $\tilde{\theta}^{NEW}$ 的获得。根据式(20)，以 \tilde{P} 表示当前参数格点的维数，则计算 Σ 需要对 $\tilde{P} \times \tilde{P}$ 的矩阵进行求逆运算，其运算量为 $O(\tilde{P}^3)$ 。值得注意的是，通过格点修剪，所需处理的参数格点维数不断减小，从而将使得求解 Σ 的运算量不断降低。并且，由于原始信号的模型阶数 K 较小，合成这一信号所需的参数格点较少，因此初始格点往往在迭代过程中被迅速修剪，这对降低求解 Σ 的运算量十分有利。事实上，在一次迭代中，EM 算法(包括变分 EM)的最大化步只需要提供一个新的参数估计以改善当前的似然函数，而不需要保证似然函数最大化，这种实现方式依然能够保证似然函数随着迭代进行不断增大(或不变)，因此依然能够求得参数的最大似然估计，且不会破坏算法的收敛性^[10]。

基于这一原理，为获得新参数格点 $\tilde{\theta}^{NEW}$ ，式(33)中的目标函数的优化只需要部分进行，这表明因获得 $\tilde{\theta}^{NEW}$ 而引起的运算量是可以方便控制的。

根据变分 EM 算法的原理^[7]，上述算法在格点修剪操作之间能够保证似然函数不断增大(或不变)，加之格点修剪操作必然是有限次的，因此上述算法是收敛的。算法的终止条件可设置为对测量 y 的拟合残差的变化量小于预设门限或算法主循环已达到最大允许循序次数。

4 应用举例及实验验证

为检验所提算法，本节将其应用于具体问题，并将其与传统 CS 重构方法进行比较。所考虑的应用实例为谱压缩感知(spectral CS)。谱压缩感知所要解决的问题是从欠采样样本中重构频率稀疏(frequency-sparse)信号^[11]。频率稀疏信号定义如下：

$$x_n = \sum_{k=1}^K a_k e^{-j\omega_k n}, \quad n = 0, \dots, N-1 \quad (34)$$

式中 $\omega_k \in [0, 2\pi)$ 表示构成信号的频率。该信号的波形匹配词典是以频率格点 $\Omega = [\Omega_1, \dots, \Omega_P]^T$ 为参数的傅里叶词典 $D(\Omega)$ ，其定义为

$$D(\Omega) = [e(\Omega_1) \quad \dots \quad e(\Omega_P)] \quad (35)$$

式中 $e(\Omega) = [e^{-j\Omega \cdot 0}, \dots, e^{-j\Omega \cdot (N-1)}]^T$ 。根据式(33)，频率格点的更新通过式(36)实现：

$$\Omega^{NEW} = \arg \min_{\Omega} \left\{ \|y - \Phi D(\Omega) \mu\|_2^2 + \text{trace}(D^H(\Omega) \Phi^H \Phi D(\Omega) \Sigma) \right\} \quad (36)$$

采用牛顿法优化式(36)。以 $H(\Omega)$ 表示目标函数，即

$$H(\Omega) = \|y - \Phi D(\Omega) \mu\|_2^2 + \text{trace}(D^H(\Omega) \Phi^H \Phi D(\Omega) \Sigma) \quad (37)$$

根据 3.4 节论述，为尽可能降低因获得 Ω^{NEW} 而引起的运算量，优化 $H(\Omega)$ 时只进行一次数值迭代。设 Ω^{OLD} 为当前频率格点，则根据牛顿法原理，有

$$\Omega^{NEW} = \Omega^{OLD} - \nabla^2 H(\Omega^{OLD})^{-1} \nabla H(\Omega^{OLD}) \quad (38)$$

式中 $\nabla H(\Omega^{OLD})$ 与 $\nabla^2 H(\Omega^{OLD})$ 分别表示 $H(\Omega)$ 在 Ω^{OLD} 处的梯度与 Hessian 矩阵。经过推导，可将两者表示如下：

$$\nabla H(\Omega^{OLD}) = -2 \text{Im}(Q^T C_1 \Phi^T u^*) - 2 \text{Im}(\text{diag}(R_1 \Sigma)) \quad (39)$$

$$\begin{aligned} \nabla^2 H(\Omega^{OLD}) &= 2 \text{Re}(Q^H C_1 \Phi^H \Phi C_1 Q + S \odot \Sigma^*) \\ &\quad + 2 \text{diag}(\text{Re}(Q^H C_2 \Phi^H u - \text{diag}(R_2 \Sigma))) \end{aligned} \quad (40)$$

式中

$$u = y - \Phi D(\Omega^{OLD}) \mu \quad (41)$$

$$\mathbf{Q} = \mathbf{D}(\boldsymbol{\Omega}^{\text{OLD}})\text{diag}(\boldsymbol{\mu}) \quad (42)$$

$$\mathbf{C}_1 = \text{diag}([0, \dots, N-1]^T) \quad (43)$$

$$\mathbf{C}_2 = \text{diag}([0^2, \dots, (N-1)^2]^T) \quad (44)$$

$$\mathbf{R}_1 = \mathbf{D}^H(\boldsymbol{\Omega}^{\text{OLD}})\mathbf{C}_1\boldsymbol{\Phi}^H\boldsymbol{\Phi}\mathbf{D}(\boldsymbol{\Omega}^{\text{OLD}}) \quad (45)$$

$$\mathbf{S} = \mathbf{D}^H(\boldsymbol{\Omega}^{\text{OLD}})\mathbf{C}_1\boldsymbol{\Phi}^H\boldsymbol{\Phi}\mathbf{C}_1\mathbf{D}(\boldsymbol{\Omega}^{\text{OLD}}) \quad (46)$$

$$\mathbf{R}_2 = \mathbf{D}^H(\boldsymbol{\Omega}^{\text{OLD}})\mathbf{C}_2\boldsymbol{\Phi}^H\boldsymbol{\Phi}\mathbf{D}(\boldsymbol{\Omega}^{\text{OLD}}) \quad (47)$$

式中 $\text{Re}(\cdot)$ 与 $\text{Im}(\cdot)$ 分别表示取实部与取虚部操作; 对任意向量 \mathbf{v} , $\text{diag}(\mathbf{v})$ 表示以 \mathbf{v} 为对角线元素的对角矩阵; 对任意方阵 \mathbf{V} , $\text{diag}(\mathbf{V})$ 表示 \mathbf{V} 的对角线元素所构成的向量; $(\cdot)^*$ 表示复共轭; \odot 表示矩阵的 Hadamard 积(即对应元素相乘)。

下面给出应用所提算法解决谱压缩感知问题的实验结果。在第 1 个实验中, 原始信号 \mathbf{x} 为包含 $8(K=8)$ 个分量频率的 $512(N=512)$ 点长信号, 其分量频率随机分布于 $[0, 2\pi)$, 复幅度随机分布在单位圆上。测量矩阵 $\boldsymbol{\Phi}$ 的行维数 $M=64$, $\boldsymbol{\Phi}$ 的各元素按照高斯分布 $\mathcal{N}(0, 1)$ 独立产生, 且将其各行的 ℓ_2 范数归一化。假设测量噪声 $\varepsilon=0$ 。按上述条件产生测量 \mathbf{y} 后, 分别采用基追踪(Basis Pursuit, BP)^[12]、稀疏贝叶斯学习(SBL)^[8]、文献[11]中所提出的基于迭代硬门限(Iterative Hard Thresholding, IHT)与线谱信号估计的谱压缩感知算法以及本文所提方法从 \mathbf{y} 中重构 \mathbf{x} 。具体地, 采用谱投影梯度(Spectral Projected Gradient, SPGL1)算法^[13]求解 BP 问题, 采用文献[14]所提出的递增 SBL 算法实现快速稀疏贝叶斯学习。对 BP 与 SBL 方法, 若将信号的稀疏表示词典设置为 DFT(Discrete Fourier Transform) 正交基, 则算法的信号重构性能将严重恶化^[4, 11]。为减轻性能恶化程度, 文献[15]建议采用过采样 DFT 作为信号的稀疏表示词典。设过采样倍数为 q , 过采样 DFT 是一个 $N \times (q \times N)$ 的矩阵, 其第 k 列($k=0, 1, \dots, qN-1$)为 $d_k(n) = e^{-j2\pi kn/(qN)}$, $n=0, 1, \dots, N-1$ 。本实验以 $q=2$ 的过采样 DFT 作为 BP 与 SBL 方法所采用的稀疏表示词典。在第 3 种方法中, 采用 ESPRIT 算法进行线谱信号估计, 按照文献[11]的符号习惯, 将这一方法简记为 SIHT-ESPRIT。在本文所提算法中, 设置初始频率格点为 $\boldsymbol{\Omega}^{\text{init}} = [0, \dots, N-1]^T \cdot 2\pi/N$, 取 $\alpha_{\text{Th}} = 100$ 。为定量衡量信号重构性能, 采用定量指标“重构信噪比(Reconstruction Signal-to-noise Ratio, RSNR)”来表征信号重构精度, 该指标定义如下:

$$\text{RSNR}(\mathbf{x}, \hat{\mathbf{x}}) = 20 \cdot \lg\left(\frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}\right) \quad (48)$$

式中 $\hat{\mathbf{x}}$ 表示 \mathbf{x} 的重构信号。

图 1 给出了原始信号以及采用不同方法重构所得信号的幅度频谱, 并列出了各方法所对应的重构信噪比。可以看出, BP 与 SBL 方法能够大致恢复出原始信号的真实频率分量, 但同时所重构信号包含大量虚假频率, 重构信噪比不高, 这表明传统的基于固定词典的 CS 重构方法不能很好适应信号频率分量偏离预设格点的非理想情况。事实上, 当信号频率分量偏离预设格点时, 采用固定词典对信号进行表示时需要利用多个格点实现对一个频率分量的拟合, 因此待重构信号相对预设词典不再具有理想的稀疏性, 这将对基于固定词典的传统 CS 重构方法的性能产生影响。不同于 BP 和 SBL, SIHT-ESPRIT 方法不是基于过冗余词典来实现信号频率分量的拟合, 而是通过 ESPRIT 算法从当前信号估计中提取信号频率。由图 1(c)可以看出, 该方法较好克服了 BP 与 SBL 方法的虚假频率问题, 但可能存在较严重的频率估计误差甚至遗漏, 从而导致信号重构精度较低。由图 1(d)看出, 与上述 3 种方法相比, 本文方法克服了这些方法的缺陷, 能够以很高精度重构原始信号, 表明本文方法能够在信号重构过程中有效优化信号的稀疏表示词典, 从而实现信号的高精度重构。

注意到上述实验中信号的测量矩阵具有随机性, 下面通过多次实验检验本文所提算法对信号测量模式的鲁棒性。为此, 采用与上述实验相同的原始信号与测量噪声设置, 进行 100 次独立实验。在每次实验中, 按照上述实验描述产生一个新的测量矩阵, 进而得到信号测量, 然后利用本文算法进行信号重构。图 2 给出了所有实验对应的重构信噪比的直方图。可以看出, 最低重构信噪比依然达到约 103 dB(对应相对误差在 10^{-6} 量级), 这说明本文算法对信号的高精度重构并不依赖于特定的测量矩阵, 它对测量矩阵的变化具有鲁棒性。

下面通过蒙特卡罗仿真检验本文算法对测量噪声的鲁棒性。实验参数设置为: $N=128$, $K=4$, $M=30$, 所考察的峰值信噪比(Peak Signal-to-noise Ratio, PSNR)的范围为 8–45 dB。峰值信噪比的定义为

$$\text{PSNR} = 20 \cdot \lg\left(\frac{\min\{|a_1|, \dots, |a_K|\}}{\sigma}\right) \quad (49)$$

式(49)中 σ 表示测量噪声的标准差。对每一个 PSNR, 进行 1000 次独立实验。在每次实验中, 按照第 1 个实验所述方法产生原始信号与测量矩阵, 按照当前 PSNR 产生复高斯白噪声, 进而按式(5)产生信号的含噪测量样本。所采用的信号重构方法与第 1 个实验相同, 有所不同的是, 基追踪方法转化为基追踪降噪(Basis Pursuit DeNoising, BPDN)

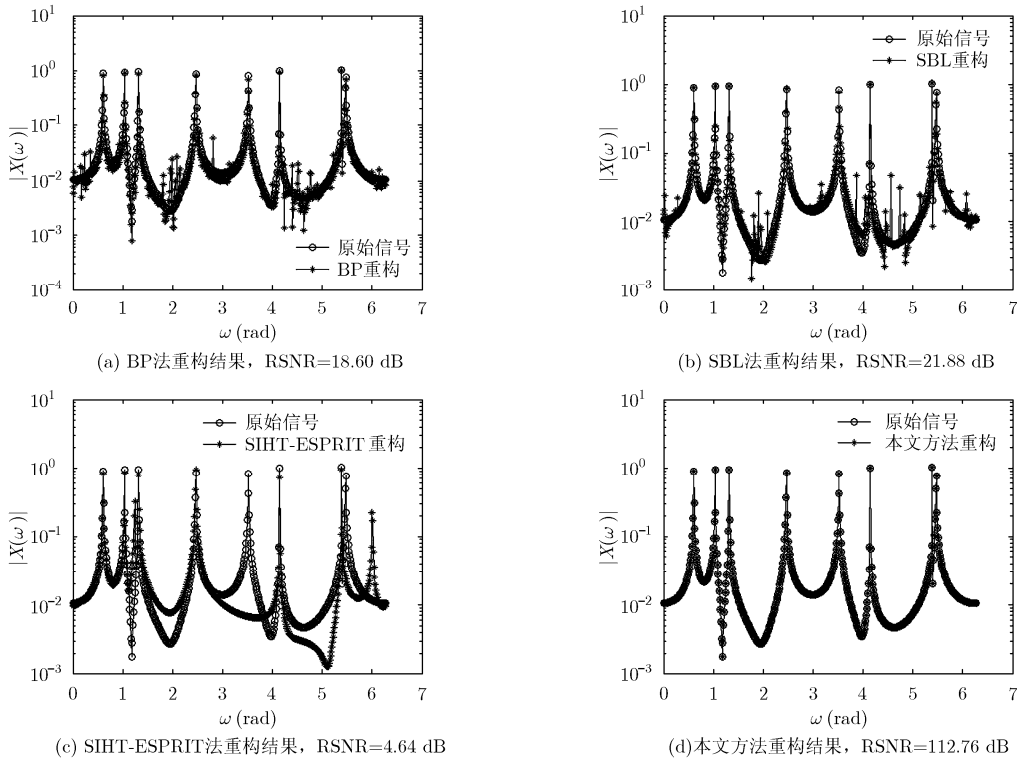


图 1 原始信号与采用不同方法重构信号的幅度谱及重构信噪比

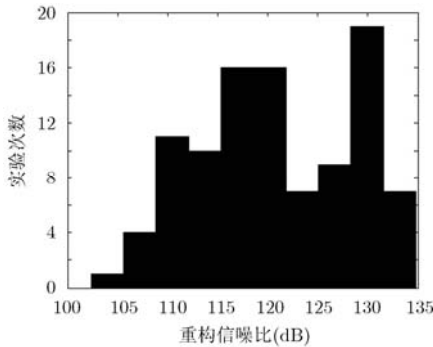


图 2 多次重构实验对应重构信噪比的直方图

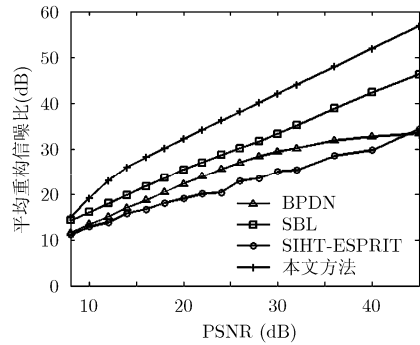


图 3 测量噪声对算法重构精度的影响

以适应测量样本被噪声污染的现实。对本文算法，初始频率格点与 α_{Th} 的设置与前面实验相同。对 BPDN 与 SBL 方法，信号的稀疏表示词典设置为过采样倍数 $q = 4$ 的过采样 DFT。需要注意的是，BPDN 方法需要估计测量噪声功率以限定测量拟合残差上限 ϵ^2 。当复测量噪声 ϵ 的方差为 σ^2 时， $\|\epsilon\|_2^2$ 是一个 χ^2 随机变量，其均值为 $M\sigma^2$ ，标准差为 $\sqrt{M}\sigma^2$ 。根据文献 [16] 中论述，本实验设定 $\epsilon^2 = (M + 2\sqrt{M})\sigma^2$ 。另外，虽然传统 SBL 算法能够在稀疏恢复过程中同时估计噪声方差，但递增 SBL 算法并不能实现噪声方差的可靠估计^[17]。基于以上考虑，本实验在使用 BPDN 与 SBL 算法时，均将噪声方差设置为真值。

图 3 示出了不同方法的平均重构信噪比随测量

噪声水平的变化曲线。可以看出，本文方法的重构精度随着峰值信噪比的增加而近似线性递增，这说明该方法的重构性能对测量噪声具有鲁棒性。与其它 3 种方法相比，本文算法显然具有更强的抑制噪声能力。值得注意的是，如前所述，本实验利用 BPDN 与 SBL 进行信号重构时，均将噪声方差设定为真实值。实际中，噪声方差往往是未知的，对其估计必然存在一定误差。因此，这两种方法对实际问题的重构性能将进一步恶化。

5 结束语

在 CS 理论中，设定合适的稀疏表示词典对信号重构至关重要，但这一词典往往难以事先确定。针对这一问题，本文提出一种基于动态词典的 CS

重构方法。该方法将稀疏表示词典建模为参数化词典,通过迭代地优化词典参数,在信号重构过程中对词典进行动态调整,从而不断减轻当前词典与真实稀疏表示词典之间的失配。为保证稀疏恢复与词典调整同步进行,该方法利用变分EM算法迭代地执行一个两步骤流程,即给定词典时恢复稀疏系数与已知稀疏系数时更新词典。文中给出了所提方法在谱压缩感知领域的应用实例,实验结果表明该方法能够取得良好的重构效果。

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] 戴琼海,付长军,季向阳. 压缩感知研究[J]. *计算机学报*, 2011, 34(3): 425-434.
Dai Qiong-hai, Fu Chang-jun, and Ji Xiang-yang. Research on compressed sensing[J]. *Chinese Journal of Computers*, 2011, 34(3): 425-434.
 - [3] 焦李成,杨淑媛,刘芳,等. 压缩感知回顾与展望[J]. *电子学报*, 2011, 39(7): 1651-1662.
Jiao Li-cheng, Yang Shu-yuan, Liu Fang, et al. Development and prospect of compressive sensing[J]. *Acta Electronica Sinica*, 2011, 39(7): 1651-1662.
 - [4] Chi Y, Scharf L L, Pezeshki A, et al. Sensitivity to basis mismatch in compressed sensing[J]. *IEEE Transactions on Signal Processing*, 2011, 59(5): 2182-2195.
 - [5] Moon T K. The expectation-maximization algorithm[J]. *IEEE Signal Processing Magazine*, 1996, 13(6): 47-60.
 - [6] Smídl V and Quinn A. *The Variational Bayes Method in Signal Processing*[M]. New York: Springer-Verlag, 2005.
 - [7] Tzikas D G, Likas A C, and Galatsanos N P. The variational approximation for Bayesian inference[J]. *IEEE Signal Processing Magazine*, 2008, 25(6): 131-146.
 - [8] Tipping M E. Sparse Bayesian learning and the relevance vector machine[J]. *Journal of Machine Learning Research*, 2001, 1: 211-244.
 - [9] Austin C D, Ertin E, Ash J N, et al. On the relation between sparse sampling and parametric estimation[C]. *IEEE 13th Digital Signal Processing Workshop and Workshop on Signal Processing Education*, Hilton Marco Island Beach Resort, Florida, 2009: 387-392.
 - [10] Neal R M and Hinton G E. *A View of the EM Algorithm that Justifies Incremental, Sparse and other Variants*[M]. Cambridge: MA, MIT Press, 1998: 355-368.
 - [11] Duarte M F and Baraniuk R G. Spectral compressive sensing[J]. *submitted to Applied and Computational Harmonic Analysis*, 2011.
 - [12] Chen S S, Donoho D L, and Saunders M A. Atomic decomposition by basis pursuit[J]. *SIAM Journal on Scientific Computing*, 1998, 20(1): 33-61.
 - [13] Van den Berg E and Friedlander M P. Probing the Pareto frontier for basis pursuit solutions[J]. *SIAM Journal on Scientific Computing*, 2009, 31(2): 890-912.
 - [14] Tipping M and Faul A. Fast marginal likelihood maximisation for sparse Bayesian models[C]. *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Key West, Florida, 2003: 1-8.
 - [15] Candès E J, Eldar Y C, Needell D, et al. Compressed sensing with coherent and redundant dictionaries[J]. *Applied and Computational Harmonic Analysis*, 2011, 31(1): 59-73.
 - [16] Candès E J, Romberg J K, and Tao T. Stable signal recovery from incomplete and inaccurate measurements[J]. *Communications on Pure and Applied Mathematics*, 2006, 59(8): 1207-1223.
 - [17] 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.
- 胡磊: 男, 1985年生, 博士生, 研究方向为压缩感知、雷达信号处理与自动目标识别。
周剑雄: 女, 1977年生, 副教授, 研究方向为目标特性、压缩感知。
石志广: 男, 1975年生, 讲师, 研究方向为压缩感知、统计分析。
付强: 男, 1962年生, 教授, 研究方向为目标特性、雷达系统设计与自动目标识别。