



基于 AMP 的 $L_{1/2}$ 正则化方法

张会¹, 张海^{1,2*}

1. 西北大学数学学院, 西安 710069

2. 中国科学院数学与系统科学院应用数学所, 北京 100190

* 通信作者. E-mail: zhanghai@nwu.edu.cn

收稿日期: 2016-06-14; 接受日期: 2016-07-08; 网络出版日期: 2016-12-02

国家自然科学基金 (批准号: 11571011) 资助项目

摘要 本文研究基于 AMP 的 $L_{1/2}$ 正则化方法, 采用 Belief Propagation 算法的思想构造了基于 AMP 改进的 Half 阈值迭代算法, 并证明所提算法至多需要有限步就能精确估计稀疏向量. 通过稀疏信号重建实验, 我们验证了几种基于 AMP 的阈值迭代求解算法的非凸正则化方法具有强的信号重建和相变能力.

关键词 $L_{1/2}$ 正则化, Half 阈值迭代算法, AMP 算法, 变量选择, 稀疏

1 引言

随着信息技术的进步, 移动互联网的普及, 产生了海量数据, 如何从海量和高维数据中进行特征提取和变量选择是我们面临的主要问题. 经典地, AIC, BIC 及 C_p [1~3] 作为几种有效的方法被用于变量选择的研究. 但是, 其求解往往对应于一个 NP 难问题, 从而不适用于海量数据分析. 因此, 如何构造处理海量与高维数据可行的方法一直以来都是数据分析研究的重点. 基于稀疏先验, 近年来发展起来的 Lasso [4~8]、SCAD [9]、MCP [10]、 $L_{1/2}$ 正则化 [11] 方法为上述问题求解提供了可能. 而关于上述正则化的算法一直是研究热点之一. 如何设计出快速和高效的实现算法, 使其更适用于海量数据处理是本文关注的问题.

上述问题可表述为如下线性模型研究:

$$y = As + \epsilon, \quad (1.1)$$

其中 y 为观测值, $y \in \mathbb{R}^p$, A 为观测矩阵, $A \in \mathbb{R}^{p \times N}$, s 是未知参数, $s \in \mathbb{R}^N$, ϵ 是随机误差, $\epsilon \in \mathbb{R}^p$. 这里假定真实模型包含 k 个变量, $k \ll N$, 我们希望从已知的 (y, A) 中选出与上述线性模型相关的少数变量, 即对稀疏线性模型进行变量选择.

Tibshirani [4]、Donoho 等 [5~7] 和 Candes 等 [8] 分别开展了上述问题的研究, 不同之处在于, Tibshirani 从变量选择角度研究, 而 Donoho 和 Candes 等是从信号的稀疏重建角度研究问题, 我们将

引用格式: 张会, 张海. 基于 AMP 的 $L_{1/2}$ 正则化方法. 中国科学: 信息科学, 2017, 47: 58–72, doi: 10.1360/N112016-00097

Zhang H, Zhang H. Approximate message passing algorithm for $L_{1/2}$ regularization (in Chinese). Sci Sin Inform, 2017, 47: 58–72, doi: 10.1360/N112016-00097

这两种方法不加区别统一称为 L_1 正则化方法. 同期, 许多学者开展了非凸正则子的研究. 这其中包括 Fan 和 Li^[9] 提出的 SCAD 正则子、Zhang^[10] 提出的 MCP 正则子、Xu 等^[11] 提出的 $L_{\frac{1}{2}}$ 正则子等非凸正则化方法.

特别地, $L_{\frac{1}{2}}$ 正则化形式如下:

$$\min_s \|y - As\|^2 + \lambda \|s\|_{\frac{1}{2}}, \quad (1.2)$$

其中 $\|s\|_{\frac{1}{2}} = \sum_{i=1}^N |s_i|^{\frac{1}{2}}$. $L_{\frac{1}{2}}$ 正则化与 L_0 正则化相比更容易求解, 而与 L_1 正则化相比能产生更稀疏的解. 说明 $L_{\frac{1}{2}}$ 正则化具有广泛并且重要的应用价值^[12~14]. 开展 $L_{\frac{1}{2}}$ 正则化快速求解算法研究具有重要的应用价值.

目前, 求解 $L_{\frac{1}{2}}$ 正则化方法的算法包括重赋权迭代算法^[11] 和 Half^[14] 阈值迭代算法. 经典的阈值迭代算法形式如下:

$$s^{(t+1)} = \eta_t(A^T r^{(t)} + s^{(t)}), \quad (1.3)$$

$$r^{(t)} = y - As^{(t)}, \quad (1.4)$$

其中 $s^{(t+1)} \in \mathbb{R}^N$ 是当前的估计值, $r^{(t)} \in \mathbb{R}^p$ 是残差, A^T 是 A 的转置, η_t 是阈值函数. 经常地, 阈值迭代算法解的稀疏度达不到模型真实的稀疏度. 近年来, 图模型^[15] 方法的研究为其提供了可能. 本文基于图模型研究成果开展 $L_{\frac{1}{2}}$ 正则化的阈值迭代算法研究.

下面简述图模型及其相关理论. 图模型通过随机变量之间的联合概率分布表示其关系, 是研究变量关系结构的一种统计方法. 通过给出图中节点的联合概率分布, 图模型将所研究问题转化为联合概率分布和边缘概率分布的求解问题. 从 Bayes 角度出发, 稀疏正则化方法对应于一个稀疏先验的似然估计问题. 通过简单对应, 稀疏正则化方法可转化为图模型的边缘分布求解.

Belief Propagation 算法 (也称为 Sum-Product) 是一种求解边缘分布的方法, Donoho 等研究发现基于逼近 Belief Propagation 算法改进的 AMP (approximate message passing) 算法^[16] 的时间复杂度远远低于 Belief Propagation 算法, 同时通过实验验证了基于 AMP 改进的 Soft 阈值迭代算法在稀疏性与欠采样的权衡方面较原阈值迭代算法有很大的改善, 新方法在稀疏性与欠采样的平衡方面与凸优化相当, 估计稀疏向量能力有很大的提高. 该方法为 L_1 正则化的求解提供了一条新的途径.

本文开展基于 AMP 的 $L_{\frac{1}{2}}$ 正则化方法的研究, 具体内容包括, 采用 Belief Propagation 算法的思想构造基于 AMP 改进的 Half 阈值迭代算法, 通过信号重建实验和相变方法验证了算法的有效性, 并证明其理论性质. 对于其他两种重要的非凸正则化方法 (SCAD, MCP), 我们通过相变方法验证了这两种非凸正则化的基于 AMP 改进的阈值迭代算法的有效性.

2 基于 AMP 的 $L_{\frac{1}{2}}$ 正则化方法

本节首先给出求解 $L_{\frac{1}{2}}$ 正则化的 Half 阈值迭代算法, 其次给出基于 AMP 改进的 Half 阈值迭代算法, 最后证明基于 AMP 改进的 Half 阈值迭代算法至多需要有限步就能精确估计稀疏向量.

2.1 Half 阈值迭代算法

为了便于叙述, 对于 $L_{\frac{1}{2}}$ 正则化的形式 (1.2) 式, 我们把 (1.3) 式中阈值迭代函数 η_t 的自变量用 u

表示, 即 $u = A^T r + s$. 那么, 基于 $L_{\frac{1}{2}}$ 正则化的 Half^[17] 阈值迭代函数可以表示成如下形式:

$$\eta_t(u; \lambda) = \begin{cases} g(u; \lambda), & |u| > \lambda, \\ 0, & |u| \leq \lambda, \end{cases}$$

其中 $g(u; \lambda) = \frac{2}{3}u(1 + \cos(\frac{2\pi}{3} - \frac{2}{3}h(u; \lambda)))$, $h(u; \lambda) = \arccos(\frac{\sqrt{2}}{2}(\frac{\lambda}{|u|})^{\frac{2}{3}})$.

给定初始值 $s^{(0)} = \mathbf{0}$, Half 阈值迭代算法可表示为

$$s^{(t+1)} = \eta_t(A^T r^{(t)} + s^{(t)}), \quad (2.1)$$

$$r^{(t)} = y - A s^{(t)}, \quad (2.2)$$

其中 $\eta_t(\cdot)$ 是 Half 阈值函数, $s^{(t)}$ 是第 t 次迭代的估计值, $r^{(t)}$ 是第 t 次迭代的残差, A^T 是 A 的转置.

2.2 基于 AMP 改进的 Half 阈值迭代算法

对于一个向量 s , $s \in \mathbb{R}^N$, A 为测量矩阵, $A \in \mathbb{R}^{p \times N}$, $p < N$, $y = A s$, $y \in \mathbb{R}^p$. 我们的目标是从 (y, A) 中估计 s . 假定 s 是稀疏的, 测量矩阵 A 满足 $A_{ai} \in \{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}$. 其中 a, b, c, \dots 和 i, j, k, \dots 分别是集合 $\{1, 2, 3, \dots, p\}$ 和 $\{1, 2, 3, \dots, N\}$ 的元素. A_{ai} 表示矩阵 A 的第 a 行第 i 列元素. 在图模型中, a, b, c, \dots 表示因子节点, i, j, k, \dots 表示变量节点.

首先考虑变量 s_1, s_2, \dots, s_N 的联合概率分布

$$\mu(ds) = \frac{1}{Z} \prod_{i=1}^N \exp(-\beta \lambda |s_i|^{\frac{1}{2}}) \prod_{a=1}^p \exp\left\{-\frac{\beta}{2}(y_a - (A s)_a)^2\right\} ds, \quad (2.3)$$

其中 Z 是归一化常数.

下面考虑因子图 $G = (V, F, E)$, 变量节点 $V = \{1, 2, \dots, N\}$, 因子节点 $F = \{1, 2, \dots, p\}$, 因子图的边 $E = \{(i, a) : i \in \{1, 2, \dots, N\}, a \in \{1, 2, \dots, p\}\}$. 这个因子图的 Belief Propagation Message Passing (MP) 算法可表示为

$$\nu_{i \rightarrow a}^{(t+1)}(s_i) \cong \exp(-\beta \lambda |s_i|^{\frac{1}{2}}) \prod_{b \neq a} \hat{\nu}_{b \rightarrow i}^{(t)}, \quad (2.4)$$

$$\hat{\nu}_{a \rightarrow i}^{(t)}(s_i) \cong \int \exp\left\{-\frac{\beta}{2}(y_a - (A s)_a)^2\right\} \prod_{j \neq i} d\nu_{j \rightarrow a}^{(t)}(s_j), \quad (2.5)$$

上标 t 表示迭代次数.

引理 2.2.1 ([17]) $\nu_{j \rightarrow a}^{(t)}$ 的分布的均值和方差分别记为 $x_{j \rightarrow a}^{(t)}$ 和 $\frac{\tau_{j \rightarrow a}^{(t)}}{\beta}$, 假定 $\int |s_j|^3 d\nu_{j \rightarrow a}^{(t)}(s_j) \leq C_t$, 则存在一个常数 C'_t 使得下面式子成立,

$$\sup_{s_i \in \mathbb{R}} |\hat{\nu}_{a \rightarrow i}^{(t)}(s_i) - \hat{\phi}_{a \rightarrow i}^{(t)}(s_i)| \leq \frac{C'_t}{N(\hat{\tau}_{a \rightarrow i}^{(t)})^{\frac{3}{2}}}, \quad (2.6)$$

其中

$$\hat{\phi}_{a \rightarrow i}^{(t)} \equiv \sqrt{\frac{\beta A_{ai}^2}{2\pi(1 + \hat{\tau}_{a \rightarrow i}^{(t)})}} \exp\left\{-\frac{\beta}{2(1 + \hat{\tau}_{a \rightarrow i}^{(t)})}(A_{ai}s_i - r_{a \rightarrow i}^{(t)})^2\right\},$$

$$r_{a \rightarrow i}^{(t)} \equiv y_a - \sum_{j \neq i} A_{aj}x_{j \rightarrow a}^{(t)}, \quad \hat{\tau}_{a \rightarrow i}^{(t)} \equiv \sum_{j \neq i} A_{aj}^2 \tau_{j \rightarrow a}^{(t)}.$$

接下来, 考虑计算 $\nu_{i \rightarrow a}^{(t+1)}(s_i)$ 的均值和方差. 为了便于叙述, 首先介绍一个密度族:

$$f_\beta(s; x, b) \equiv \frac{1}{Z_\beta(x, b)} \exp \left\{ -\beta |s|^{\frac{1}{2}} - \frac{\beta}{2b} (s-x)^2 \right\}. \quad (2.7)$$

定义此密度族的均值和方差 (Z 的密度函数为 $f_\beta(\cdot; x, b)$) 分别为

$$F_\beta(x; b) \equiv \mathbb{E}_{f_\beta(\cdot; x, b)}(Z), \quad G_\beta(x; b) \equiv \text{Var}_{f_\beta(\cdot; x, b)}(Z). \quad (2.8)$$

引理 2.2.2 假定在第 t 步迭代时, 因子节点到变量节点的信息满足 $\hat{\nu}_{a \rightarrow i}^{(t)} = \hat{\phi}_{a \rightarrow i}^{(t)}$, 参数 $\hat{\tau}_{a \rightarrow i}^{(t)} = \hat{\tau}^{(t)}$. 则在下一步迭代有下面结论成立:

$$\nu_{i \rightarrow a}^{(t)}(s_i) = \phi_{i \rightarrow a}^{(t)} \left\{ 1 + O \left(\frac{s_i^2}{p} \right) \right\}, \quad \phi_{i \rightarrow a}^{(t)} \equiv \lambda^2 f_\beta \left(\lambda^2 s_i; \lambda^2 \sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)}, \lambda^4 (1 + \hat{\tau}^{(t)}) \right). \quad (2.9)$$

特别地, $\nu_{i \rightarrow a}^{(t)}$ 的均值和方差为

$$x_{i \rightarrow a}^{(t+1)}(s_i) = \frac{1}{\lambda^2} F_\beta \left(\lambda^2 \sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^4 (1 + \hat{\tau}^{(t)}) \right), \quad \tau_{i \rightarrow a}^{(t)} = \frac{\beta}{\lambda^4} G_\beta \left(\lambda^2 \sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^4 (1 + \hat{\tau}^{(t)}) \right). \quad (2.10)$$

证明 由方程 (2.4) 可以得到

$$\begin{aligned} \nu_{i \rightarrow a}^{(t)}(s_i) &\cong \exp \left\{ -\beta \lambda |s_i|^{\frac{1}{2}} \right\} \prod_{b \neq a} \hat{\nu}_{b \rightarrow i}^{(t)} \\ &\cong \exp \left\{ -\beta \lambda |s_i|^{\frac{1}{2}} - \sum_{b \neq a} \frac{\beta}{2(1 + \hat{\tau}^{(t)})} (A_{bi} s_i - r_{b \rightarrow i}^{(t)})^2 \right\} \\ &\cong \exp \left\{ -\beta \lambda |s_i|^{\frac{1}{2}} - \frac{\beta}{2(1 + \hat{\tau}^{(t)})} \left(\frac{p-1}{p} s_i^2 - 2s_i \sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)} \right) \right\} \\ &\cong \exp \left\{ -\beta |\lambda^2 s_i|^{\frac{1}{2}} - \frac{\beta}{2\lambda^4 (1 + \hat{\tau}^{(t)})} \left(\lambda^2 s_i - \lambda^2 \sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)} \right)^2 + \frac{\beta}{2(\hat{\tau}^{(t)} + 1)} \frac{s_i^2}{p} \right\}, \end{aligned}$$

上面第 3 个等价式说明 $\nu_{i \rightarrow a}^{(t)}(s_i)$ 和 $\phi_{i \rightarrow a}^{(t)}$ 的误差的数量级为 $O(\frac{s_i^2}{p})$, 从而可得结论. 由定义可以直接计算得到 F_β 和 G_β , 从而引理 2.2.2 得证.

综合以上讨论, 并且假定 $\hat{\tau}_{a \rightarrow i}^{(t)}$ 与边 (a, i) 独立, 即满足 $\hat{\tau}_{a \rightarrow i}^{(t)} = \hat{\tau}^{(t)}$, 则可以得到如下 MP 算法:

$$x_{i \rightarrow a}^{(t+1)}(s_i) = \frac{1}{\lambda^2} F_\beta \left(\lambda^2 \sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^4 (1 + \hat{\tau}^{(t)}) \right), \quad r_{a \rightarrow i}^{(t)} \equiv y_a - \sum_{j \neq i} A_{aj} x_{j \rightarrow a}^{(t)}, \quad (2.11)$$

$$\tau^{(t+1)} = \frac{\beta}{\lambda^4 p} \sum_{i=1}^N G_\beta \left(\lambda^2 \sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^4 (1 + \hat{\tau}^{(t)}) \right). \quad (2.12)$$

回顾 (1.2) 和 (2.3) 式, 可以得出当 $\beta \rightarrow \infty$ 时, 最大化 (2.3) 式的解 $F_\beta(x; b)$ 对应于 (1.2) 式的解 $\eta(x; b)$, 即

$$F_\beta(x; b) \rightarrow \eta(x; b), \quad \beta \rightarrow \infty. \quad (2.13)$$

通过下面的引理说明这种对应关系.

引理 2.2.3 ([17]) 对任意有界的 x, b , 有下面结论成立:

$$\lim_{\beta \rightarrow \infty} F_\beta(x; \beta) = \eta(x; b), \quad \lim_{\beta \rightarrow \infty} \beta G_\beta(x; \beta) = b\eta'(x; b). \quad (2.14)$$

综合引理 2.2.1~2.2.3, 并且结合 Half 阈值函数的不变性 $\eta(ax; ab) = a\eta(x; b)(\forall a > 0)$. 可以得到 MP 算法:

$$x_{i \rightarrow a}^{(t+1)} = \eta_t \left(\sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^2 (1 + \hat{\tau}^{(t)}) \right), \quad r_{a \rightarrow i}^{(t)} \equiv y_a - \sum_{j \neq i} A_{aj} x_{j \rightarrow a}^{(t)}, \quad (2.15)$$

$$\hat{\tau}^{(t+1)} = \frac{1 + \hat{\tau}^{(t)}}{N\delta} \sum_{i=1}^N \eta'_t \left(\sum_b A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^2 (1 + \hat{\tau}^{(t)}) \right). \quad (2.16)$$

令 $\lambda^2 \hat{\tau}^{(t)} = \gamma^{(t)}$, 则可以得到 MP 算法的新形式:

$$x_{i \rightarrow a}^{(t+1)} = \eta_t \left(\sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^2 + \gamma^{(t)} \right), \quad r_{a \rightarrow i}^{(t)} \equiv y_a - \sum_{j \neq i} A_{aj} x_{j \rightarrow a}^{(t)}, \quad (2.17)$$

$$\gamma^{(t+1)} = \frac{\lambda^2 + \gamma^{(t)}}{N\delta} \sum_{i=1}^N \eta'_t \left(\sum_b A_{bi} r_{b \rightarrow i}^{(t)}; \lambda^2 + \gamma^{(t)} \right). \quad (2.18)$$

MP 算法 (2.17) 和 (2.18) 每一步迭代需要更新 $2pN$ 个信息, 计算复杂度仍然很大, 本节的目的是简化 MP 算法的更新规则, 我们希望这种小的修正可以大大降低计算复杂度.

假定信息可以用下面方式逼近:

$$x_{i \rightarrow a}^{(t)} = x_i^{(t)} + \delta x_{i \rightarrow a}^{(t)} + O\left(\frac{1}{N}\right), \quad (2.19)$$

$$r_{i \rightarrow a}^{(t)} = r_i^{(t)} + \delta r_{i \rightarrow a}^{(t)} + O\left(\frac{1}{N}\right), \quad (2.20)$$

其中 $\delta x_{i \rightarrow a}^{(t)}, \delta r_{i \rightarrow a}^{(t)} = O(\frac{1}{\sqrt{N}})$. 接下来, 考虑一般的 MP 算法形式:

$$x_{i \rightarrow a}^{(t+1)} = \eta_t \left(\sum_{b \neq a} A_{bi} r_{b \rightarrow i}^{(t)} \right), \quad r_{a \rightarrow i}^{(t)} \equiv y_a - \sum_{j \neq i} A_{aj} x_{j \rightarrow a}^{(t)}, \quad (2.21)$$

其中 $\{\eta_t(\cdot)\}_{t \rightarrow N}$ 是一个非线性可微函数序列, 并且导数有界, 则有下面结论成立.

引理 2.2.4 ([17]) 假定 MP 算法 (2.21) 的逼近式 (2.19) 和 (2.20) 成立, 则 $x_i^{(t)}$ 和 $r_i^{(t)}$ 满足下面方程:

$$x_i^{(t+1)} = \eta_t \left(\sum_a A_{ia} r_a^{(t)} + x_i^{(t)} \right) + o_N(1), \quad (2.22)$$

$$r_a^{(t)} = y_a - \sum_j A_{aj} x_j^{(t)} + \frac{1}{\delta} r_a^{(t-1)} \langle \eta'_{t-1}(A^T r^{(t-1)} + x^{(t-1)}) \rangle + o_N(1), \quad (2.23)$$

当 $p, N \rightarrow \infty$ 时, $o_N(1) \rightarrow 0$.

由引理 2.2.4 可以得到基于 AMP 改进的 Half 阈值迭代算法, 其中 $\delta = \frac{p}{N}$, 对于一个向量 $u = (u_1, u_2, \dots, u_N)$, $\langle u \rangle = \sum_{i=1}^N u_i/N$, $\eta'_t(s) = \frac{\partial}{\partial s} \eta_t(s)$,

$$x^{(t+1)} = \eta_t(x^{(t)} + A^T r^{(t)}; \lambda^2 + \gamma^{(t)}), \quad (2.24)$$

$$r^{(t)} = y - Ax^{(t)} + \frac{1}{\delta} r^{(t-1)} \langle \eta'_{t-1}(x^{(t-1)} + A^T r^{(t-1)}; \lambda^2 + \gamma^{(t-1)}) \rangle, \quad (2.25)$$

$$\gamma^{(t+1)} = \frac{\lambda^2 + \gamma^{(t)}}{\delta} \langle \eta'_t(x^{(t)} + A^T r^{(t)}; \lambda^2 + \gamma^{(t)}) \rangle. \quad (2.26)$$

2.3 基于 AMP 改进的 Half 阈值迭代算法的理论分析

为方便研究该算法的理论性质, 我们给出一些相关记号. 首先, 记 A_i 为 A 的第 i 列向量, s_i 为 s 的第 i 个元素, J 为指标集. 定义

$$A_J = \{A_i \mid i \in J\} \quad s_J = \{s_i \mid i \in J\}, \quad (2.27)$$

$s^{(t)}$ 的主动集 (若 s 的元素 s_j 通过阈值, 则把 s_j 放入集合 I 中 (即 $j \in I$), 集合 I 就称为主动集) 记为 $I^{(t)}$, s 的主动集记为 I^* , k 为 s 的稀疏度 (即 s 至多有 k 个非零的元素). A 的相关性记为

$$\mu = \max_{\{i,j:1 \leq i,j \leq N, i \neq j\}} |\langle A_i, A_j \rangle|. \quad (2.28)$$

不失一般性, 假定 $\|A_j\|_2 = 1$, $j = 1, 2, \dots, N$. s 的元素按绝对值大小递减排列 (即 s 的前 k 个元素非零, 且 $|s_1| \geq |s_2| \geq \dots \geq |s_k| \geq 0$).

接下来, 定义两个变量,

$$z^{(t+1)} = (I - A^T A)s^{(t)} + A^T A s + \frac{1}{\delta} A^T r^{(t-1)} \langle \eta'_{t-1}(A^T r^{(t-1)} + s^{(t-1)}) \rangle, \quad (2.29)$$

$$w^{(t)} = s - s^{(t)}, \quad (2.30)$$

这里 s 是最优值, $s^{(t)}$ 是第 t 步的估计值, $z^{(t+1)}$ 和 $w^{(t)}$ 的第 j 个元素记为 $z_j^{(t+1)}$ 和 $w_j^{(t)}$.

最后, 讨论 Half 阈值迭代函数的导数,

$$\frac{1}{\delta} \langle \eta'_{t-1}(A^T r^{(t-1)} + s^{(t-1)}) \rangle \leq \frac{N}{p} \cdot \frac{k}{N} \cdot \frac{5}{3} = \frac{k}{p} \cdot \frac{5}{3}, \quad (2.31)$$

记 $e = \frac{k}{p} \cdot \frac{5}{3}$, 一般情况下, k 远小于 p , 所以 $e < 1$ 成立.

定理 2.3.1 假定 $k\mu < \varepsilon$ (ε 为正数), 且对任意的 $1 \leq i < k$, 都有 $\frac{|s_i|}{|s_{i+1}|} < \alpha_i$ (α_i 为正数) 成立, 则基于 AMP 改进的 Half 阈值迭代算法至多需要 $L = \sum_{i=1}^{k-1} l_i + 1$ 步 (l_i 为正整数) 就能估计稀疏向量 s , 并且当 $t \geq L$ 时, 有

$$I^* \subset I^{(t)}. \quad (2.32)$$

注 2.3.1 定理 2.3.1 表明基于 AMP 改进的非凸阈值迭代算法只需有限步就可以找到正确的主动集.

下面, 给出定理 2.3.1 的证明. 迭代初始值选为 $s^{(0)} = \mathbf{0}$. 为了证明定理 2.3.1, 需要以下几个引理.

引理 2.3.1 假定 $k\mu < \frac{1}{2}$, 则 s 的最大元素 s_1 在第 1 步迭代后进入主动集中, 并且对任意的 j 有,

$$|z_j^{(1)} - s_j| < \frac{1}{2}|s_1|. \quad (2.33)$$

证明

$$\begin{aligned} |z_1^{(1)}| &\geq |s_1| - |z_1^{(1)} - s_1| = |s_1| - \left| \sum_{j \in I^* \cup I^{(0)} \setminus \{1\}} \langle A_1, A_j \rangle w_j^{(0)} \right| \\ &= |s_1| - \left| \sum_{j \in I^* \setminus \{1\}} \langle A_1, A_j \rangle s_j \right| \geq |s_1| - k\mu |s_1|, \end{aligned}$$

上式第 3 个等式成立是由于假设迭代初始值为零 (即 $s_j^{(0)} = 0$), 则就有 $I^{(0)}$ 为空集, 上式第 4 个不等式成立是由于当 $j \in I^* \setminus I^{(0)}$ 时, $w_j^{(0)} = s_j$, 且有 $|s_j| \leq |s_1|$. 这里只考虑 s 的最大元素 s_1 在第 1 步迭代后是否进入主动集中, 对于 $1 < i \leq k$ 不做讨论. 当 $i > k$ 时有,

$$\max_{\{i:k < i\}} |z^{(1)}(i)| = \max_{\{i:k < i\}} \left| \sum_{j=1}^k \langle A_i, A_j \rangle s_j \right| \leq k\mu |s_1|,$$

由于假设 $k\mu < \frac{1}{2}$, 所以有 $k\mu < 1 - k\mu$, 故 $|z_1^{(1)}| > \max_{\{i:k < i\}} |z^{(1)}(i)|$, 因此, s 的最大元素 s_1 在第 1 步迭代后进入主动集中, 并且有

$$|z_j^{(1)} - s_j| = \left| \sum_{i \in I^* \setminus \{j\}} \langle A_i, A_j \rangle s_i \right| \leq k\mu |s_1| < \frac{1}{2} |s_1|.$$

引理 2.3.2 假定 $k\mu < \varepsilon_1 \leq \frac{1}{2}$, 则在第 $1+n$ 步迭代后 s_1 仍在主动集中, 并且对任意的 j 有,

$$|z_j^{(1+n)} - s_j| < \frac{1}{2} |s_1|. \quad (2.34)$$

证明 用数学归纳法证明这个引理, 首先证明当 $n = 1$ 时结论成立.

$$\begin{aligned} |z_i^{(2)} - s_i| &\leq \sum_{j \in I^{(1)} \setminus \{i\}} |\langle A_i, A_j \rangle w_j^{(1)}| + \sum_{j \in I^* \setminus I^{(1)} \cup \{i\}} |\langle A_i, A_j \rangle w_j^{(1)}| + e |[A^T r^{(0)}]_i| \\ &\leq \sum_{j \in I^{(1)} \setminus \{i\}} |\langle A_i, A_j \rangle (s_j - z_j^{(1)})| + k\mu |s_2| + e \sum_{j \in I^*} |\langle A_i, A_j \rangle s_j| \\ &\leq \frac{1}{2} k\mu |s_1| + k\mu |s_2| + ek\mu |s_1| \\ &\leq \left(\frac{3}{2} + e \right) k\mu |s_1|, \end{aligned}$$

令 $\varepsilon_1^{(1)} = \frac{1}{3+2e}$, 若 $\varepsilon_1 \leq \varepsilon_1^{(1)}$, 则

$$|z_i^{(2)} - s_i| < \left(\frac{3}{2} + e \right) \varepsilon_1 |s_1| < \frac{1}{2} |s_1|.$$

下面将证明 s 的第 1 个元素仍在主动集中. 当 $i = 1$ 时,

$$|z_1^{(2)}| \geq |s_1| - |z_1^{(2)} - s_1| > \frac{|s_1|}{2}.$$

当 $i > k$ 时,

$$|z_i^{(2)}| = |z_i^{(2)} - s_i| < \frac{1}{2} |s_1|,$$

由于 $|z_1^{(2)}| > \max_{\{i:k < i\}} |z_i^{(2)}|$, 因此, s 的第 1 个元素仍在主动集中. 从而 $n = 1$ 时结论成立.

接下来, 假定 $n < t$ 时结论成立, 证明 $n = t$ 时结论成立, 与 $n = 1$ 时证明过程类似有

$$\begin{aligned} |z_i^{(t+1)} - s_i| &\leq \sum_{j \in I^{(t)} \setminus \{i\}} |\langle A_i, A_j \rangle w_j^{(t)}| + \sum_{j \in I^* \setminus I^{(t)} \cup \{i\}} |\langle A_i, A_j \rangle w_j^{(t)}| + |e[A^T r^{(t-1)}]_i| \\ &\leq \frac{1}{2} k\mu |s_1| + k\mu |s_2| + \frac{1}{2} k\mu |s_1| \sum_{i=1}^{t-1} e^i + k\mu |s_2| \sum_{i=1}^{t-1} e^i + k\mu |s_1| e^{t-1} \\ &\leq k\mu |s_1| \left[\frac{3}{2} \sum_{i=0}^{t-1} e^i + e^{t-1} \right], \end{aligned}$$

由于 $e < 1$, 所以 $\sum_{i=0}^{t-1} e^i$ 存在, 令

$$\varepsilon_1^{(t)} = \frac{1}{3 \sum_{i=0}^{t-1} e^i + 2e^{t-1}},$$

若 $\varepsilon_1 \leq \varepsilon_1^{(t)}$, 则

$$|z_i^{(t+1)} - s_i| < \frac{1}{2} |s_1|,$$

易得

$$|z_1^{(t+1)}| \geq |s_1| - |z_1^{(t+1)} - s_1| > \frac{|s_1|}{2}.$$

当 $i > k$ 时,

$$|z_i^{(t+1)}| = |z_i^{(t+1)} - s_i| < \frac{1}{2} |s_1|,$$

由于 $|z_1^{(t+1)}| > \max_{\{i:k < i\}} |z_i^{(t+1)}|$, 因此, s 的第 1 个元素仍在主动集中, 从而第 $t+1$ 步迭代结论成立. 令

$$\varepsilon_1 = \min \left\{ \frac{1}{2}, \varepsilon_1^{(1)}, \varepsilon_1^{(2)}, \dots, \varepsilon_1^{(n)} \right\},$$

引理结论成立.

引理 2.3.3 假定 $k\mu < \varepsilon_2^{(1)} \leq \varepsilon_1$, $\frac{|s_1|}{|s_2|} < \alpha_1$ 成立, 那么再经过 l_1 步迭代后, s_2 将进入主动集中, 并且对任意的 j 有

$$|z_j^{(1+l_1+1)} - s_j| < \frac{1}{2} |s_2|. \quad (2.35)$$

证明 令 $n = l_1$, 由引理 2.3.2 可得

$$\begin{aligned} |z_i^{(l_1+1)} - s_i| &\leq \frac{1}{2} k\mu |s_1| \sum_{i=0}^{l_1-1} e^i + k\mu |s_2| \sum_{i=0}^{l_1-1} e^i + k\mu |s_1| e^{l_1-1} \\ &< \varepsilon_1 |s_2| \left[\left(\frac{1}{2} \sum_{i=0}^{l_1-1} e^i + e^{l_1-1} \right) \alpha_1 + \sum_{i=0}^{l_1-1} e^i \right], \end{aligned}$$

令

$$\alpha_1 = \frac{1 - 2\varepsilon_1 \sum_{i=0}^{l_1-1} e^i}{\varepsilon_1 \sum_{i=0}^{l_1-1} e^i + 2\varepsilon_1 e^{l_1-1}},$$

可得

$$|z_i^{(l_1+1)} - s_i| < \frac{1}{2} |s_2|,$$

同引理 2.3.2 证明, 得

$$|z_2^{(l_1+1)}| \geq |s_2| - |z_2^{(l_1+1)} - s_2| > \frac{|s_2|}{2}.$$

当 $i > k$ 时,

$$|z_i^{(l_1+1)}| = |z_i^{(l_1+1)} - s_i| < \frac{1}{2}|s_2|,$$

由于 $|z_2^{(l_1+1)}| > \max_{\{i:k < i\}} |z_i^{(l_1+1)}|$, 因此, s 的第 2 个元素在主动集中, 并且对任意的 j 有

$$\begin{aligned} |z_i^{(l_1+1+1)} - s_i| &\leq \sum_{j \in I^{(l_1+1)} \setminus \{i\}} |\langle A_i, A_j \rangle w_j^{(l_1+1)}| + \sum_{j \in I^* \setminus I^{(l_1+1)} \cup \{i\}} |\langle A_i, A_j \rangle w_j^{(l_1+1)}| + |e[A^T r^{(l_1)}]_i| \\ &\leq k\mu |z_i^{(l_1+1)} - s_i| + k\mu |s_3| + \frac{1}{2}k\mu |s_1| \sum_{i=1}^{l_1} e^i + k\mu |s_2| \sum_{i=1}^{l_1} e^i + k\mu |s_1| e^{l_1} \\ &\leq k\mu |s_2| \left[\sum_{i=0}^{l_1} e^i + \frac{1}{2} + \frac{\alpha_1}{2} \sum_{i=1}^{l_1} e^i + \alpha_1 e^{l_1} \right], \end{aligned}$$

令

$$\begin{aligned} \tilde{\varepsilon}_1 &= \frac{1}{2 \sum_{i=0}^{l_1} e^i + 1 + \alpha_1 \sum_{i=1}^{l_1} e^i + 2\alpha_1 e^{l_1}}, \\ \varepsilon_2^{(1)} &= \min\{\varepsilon_1, \tilde{\varepsilon}_1\}, \end{aligned}$$

可得若 $k\mu < \varepsilon_2^{(1)}$, 则对任意的 j ,

$$|z_j^{(1+l_1+1)} - s_j| < \frac{1}{2}|s_2|,$$

从而引理成立.

引理 2.3.4 假定 $k\mu < \varepsilon_2 \leq \varepsilon_2^{(1)}$, 则在第 $1 + l_1 + n$ 步迭代后 s_1, s_2 仍在主动集中, 并且对任意的 j 有,

$$|z_j^{(1+l_1+n)} - s_j| < \frac{1}{2}|s_2|. \quad (2.36)$$

证明 用数学归纳法证明这个引理, 首先当 $n = 1$ 时, 由引理 2.3.3 的证明可得若 $k\mu < \varepsilon_2^{(1)}$, 则有

$$|z_j^{(1+l_1+1)} - s_j| < \frac{1}{2}|s_2|.$$

下面将证明 s 的第 2 个元素仍在主动集中. 当 $i = 2$ 时,

$$|z_2^{(1+l_1+1)}| \geq |s_2| - |z_2^{(1+l_1+1)} - s_2| > \frac{|s_2|}{2}.$$

当 $i > k$ 时,

$$|z_i^{(1+l_1+1)}| = |z_i^{(1+l_1+1)} - s_i| < \frac{1}{2}|s_2|,$$

由于 $|z_2^{(1+l_1+1)}| > \max_{\{i:k < i\}} |z_i^{(1+l_1+1)}|$, 因此, s 的第 2 个元素仍在主动集中. 从而 $n = 1$ 时结论成立.

接下来, 假定 $n < t$ 时结论成立, 证明 $n = t$ 时结论成立, 与 $n = 1$ 时证明过程类似有

$$|z_i^{(t+l_1+1)} - s_i| \leq \sum_{j \in I^{(t+l_1)} \setminus \{i\}} |\langle A_i, A_j \rangle w_j^{(t+l_1)}| + \sum_{j \in I^* \setminus I^{(t+l_1)} \cup \{i\}} |\langle A_i, A_j \rangle w_j^{(t+l_1)}| + |e[A^T r^{(t+l_1-1)}]_i|$$

$$\begin{aligned} &\leq \frac{1}{2}k\mu|s_2| + k\mu|s_3| + \frac{1}{2}k\mu|s_1| \sum_{i=1}^{t+l_1-1} e^i + k\mu|s_2| \sum_{i=1}^{t+l_1-1} e^i + k\mu|s_1|e^{t+l_1-1} \\ &\leq k\mu|s_2| \left[\frac{3}{2} + \left(\frac{\alpha_1}{2} + 1 \right) \sum_{i=1}^{t+l_1-1} e^i + \alpha_1 e^{t+l_1-1} \right], \end{aligned}$$

令

$$\varepsilon_2^{(t)} = \frac{1}{[3 + (\alpha_1 + 2) \sum_{i=1}^{t+l_1-1} e^i + 2\alpha_1 e^{t+l_1-1}]},$$

若 $\varepsilon_2 \leq \varepsilon_2^{(t)}$, 则

$$|z_i^{(t+l_1+1)} - s_i| < \frac{1}{2}|s_2|,$$

易得

$$|z_2^{(t+l_1+1)}| \geq |s_2| - |z_2^{(t+l_1+1)} - s_2| > \frac{|s_2|}{2}.$$

当 $i > k$ 时,

$$|z_i^{(t+l_1+1)}| = |z_i^{(t+l_1+1)} - s_i| < \frac{1}{2}|s_2|,$$

由于 $|z_2^{(t+l_1+1)}| > \max_{\{i:k < i\}} |z_i^{(t+l_1+1)}|$, 因此, s 的第 2 个元素仍在主动集中, 从而第 $t + l_1 + 1$ 步迭代结论成立. 令 $\varepsilon_2 = \min\{\varepsilon_2^{(1)}, \varepsilon_2^{(2)}, \dots, \varepsilon_2^{(n)}\}$, 引理结论成立.

引理 2.3.5 假定 $k\mu < \varepsilon_3^{(1)} \leq \varepsilon_2$, $\frac{|s_2|}{|s_3|} < \alpha_2$ 成立, 那么再经过 l_2 步迭代后, s_3 将进入主动集中, 并且对任意的 j 有

$$|z_j^{(1+l_1+l_2+1)} - s_j| < \frac{1}{2}|s_3|. \tag{2.37}$$

证明 令 $n = l_2$, 由引理 2.3.4 可得

$$\begin{aligned} |z_i^{(l_1+l_2+1)} - s_i| &\leq \frac{1}{2}k\mu|s_2| + k\mu|s_3| + \frac{1}{2}k\mu|s_1| \sum_{i=1}^{l_2+l_1-1} e^i + k\mu|s_2| \sum_{i=1}^{l_2+l_1-1} e^i + k\mu|s_1|e^{l_2+l_1-1} \\ &< \left[\frac{\alpha_2}{2} + 1 + \alpha_2 \left(\frac{\alpha_1}{2} + 1 \right) \sum_{i=1}^{l_2+l_1-1} e^i + \alpha_1 \alpha_2 e^{l_2+l_1-1} \right] \varepsilon_2 |s_3|, \end{aligned}$$

令

$$\alpha_2 = \frac{(1 - 2\varepsilon_2)}{\varepsilon_2 [1 + (\alpha_1 + 2) \sum_{i=1}^{l_2+l_1-1} e^i + 2\alpha_1 e^{l_2+l_1-1}]},$$

可得

$$|z_i^{(l_1+l_2+1)} - s_i| < \frac{1}{2}|s_3|,$$

同引理 2.3.4 证明, 得

$$|z_3^{(l_1+l_2+1)}| \geq |s_3| - |z_3^{(l_1+l_2+1)} - s_3| > \frac{|s_3|}{2}.$$

当 $i > k$ 时,

$$|z_i^{(l_1+l_2+1)}| = |z_i^{(l_1+l_2+1)} - s_i| < \frac{1}{2}|s_3|,$$

由于 $|z_2^{(l_1+l_2+1)}| > \max_{\{i:k < i\}} |z_i^{(l_1+l_2+1)}|$, 因此, s 的第 3 个元素在主动集中, 并且对任意的 j 有

$$\begin{aligned} |z_i^{(l_1+l_2+1+1)} - s_i| &\leq \sum_{j \in I^{(l_1+l_2+1)} \setminus \{i\}} |\langle A_i, A_j \rangle w_j^{(l_1+l_2+1)}| \\ &\quad + \sum_{j \in I^* \setminus I^{(l_1+l_2+1)} \cup \{i\}} |\langle A_i, A_j \rangle w_j^{(l_1+l_2+1)}| + |e[A^T r^{(l_1+l_2)}]_i| \\ &\leq k\mu |z_i^{(l_1+l_2+1)} - s_i| + k\mu |s(4)| + \frac{1}{2} k\mu |s_1| \sum_{i=1}^{l_1+l_2} e^i + k\mu |s_2| \sum_{i=1}^{l_1+l_2} e^i + k\mu |s_1| e^{l_1+l_2} \\ &\leq k\mu |s_3| \left[\frac{1}{2} + 1 + \alpha_2 \left(\frac{\alpha_1}{2} + 1 \right) \sum_{i=1}^{l_1+l_2} e^i + \alpha_1 \alpha_2 e^{l_1+l_2} \right], \end{aligned}$$

令

$$\begin{aligned} \tilde{\varepsilon}_2 &= \frac{1}{3 + \alpha_2(\alpha_1 + 2) \sum_{i=1}^{l_1+l_2} e^i + 2\alpha_1 \alpha_2 e^{l_1+l_2}}, \\ \varepsilon_3^{(1)} &= \min\{\varepsilon_2, \tilde{\varepsilon}_2\}, \end{aligned}$$

可得若 $k\mu < \varepsilon_3^{(1)}$, 则对任意的 j ,

$$|z_j^{(1+l_1+l_2+1)} - s_j| < \frac{1}{2} |s_3|,$$

从而引理成立.

定理 2.3.1 的证明 由引理 2.3.1 ~ 2.3.5, 依此类推可以得到 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k$ ($\varepsilon_1 \geq \varepsilon_2 \geq \dots \geq \varepsilon_k$), $\alpha_1, \alpha_2, \dots, \alpha_{k-1}$, 以及 l_1, l_2, \dots, l_{k-1} . 若 $\varepsilon \leq \varepsilon_k$, 则当 $k\mu < \varepsilon$ (ε 为正数), 且对任意的 $1 \leq i < k$, $\frac{|s_i|}{|s_{i+1}|} < \alpha_i$ (α_i 为正数) 成立时, 基于 AMP 改进的 Half 阈值迭代算法至多需要 $L = \sum_{i=1}^{k-1} l_i + 1$ 步 (l_i 为正整数) 就能估计稀疏向量 s , 并且当 $t \geq L$ 时, 令 $\varepsilon = \min\{\varepsilon_k, \varepsilon_{k+1}^{(1)}, \varepsilon_{k+1}^{(2)}, \dots, \varepsilon_{k+1}^{(t-L)}\}$, 就有 $I^* \subset I^{(t)}$, 从而定理得证.

3 实验

本节将基于 AMP 改进的非凸阈值迭代算法应用于压缩感知问题的研究, 并通过相变方法研究新算法的性能.

我们将基于 AMP 改进的 Half 阈值迭代算法应用于压缩感知问题的求解. 压缩感知^[18,19], 又称为稀疏信号重构, 是研究稀疏或者可压缩信号的重建方法, 近年来在统计机器学习、图像处理和信号处理等多个领域得到广泛应用^[20,21]. 压缩感知是这样一类新的信息获取技术, 它通过信息表示的稀疏性, 将采样和压缩两种方式合并进行来获取信息, 相比于传统 Nyquist 采样的方式, 它具有更小的采样率. 压缩感知突破了传统的采样方式, 利用更加经济的形式来获取信息, 因此, 它为高复杂性信息的获取、处理及应用提供了可能性.

压缩感知本质上研究的是如下欠定线性方程组的求解问题:

$$y = As, \quad A \in \mathbb{R}^{p \times N} \quad (p \ll N), \quad (3.1)$$

其中 s 是信号, $s \in \mathbb{R}^N$, A 是测量矩阵, $A \in \mathbb{R}^{p \times N}$, y 是测量向量, $y \in \mathbb{R}^p$. 我们希望从观测数据 y 恢复稀疏未知信号 s . 其数学模型如下:

$$\min_s \|y - As\|^2 + P(s; \lambda), \quad (3.2)$$

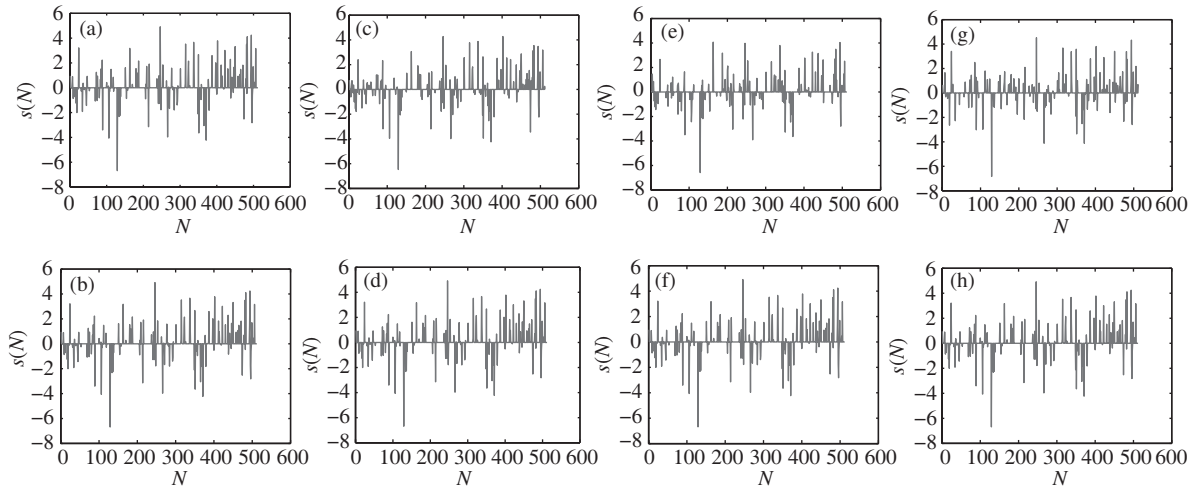


图 1 Half 阈值迭代算法与基于 AMP 改进的 Half 阈值迭代算法信号重建的实验结果

Figure 1 The signal reconstruction results of Half iterative thresholding algorithm and Half iterative thresholding algorithm based on AMP. (a) Original; (b) Half(240); (c) Half(239); (d) Half AMP(239); (e) Half(235); (f) Half AMP(235); (g) Half(233); (h) Half AMP(233)

其中 $P(s; \lambda)$ 为惩罚函数, 现有模型如基于 L_1 正则化的压缩感知^[7], 基于 $L_{\frac{1}{2}}$ 正则化的压缩感知^[11], 基于 SCAD 正则化的压缩感知^[22], 基于 MCP 正则化的压缩感知^[10] 等工作.

下面研究基于 $L_{\frac{1}{2}}$ 正则化的压缩感知问题.

实验 1 我们研究这样的信号重建问题: 假设 s 为 $N = 512$, 取含有 130 个非零窄脉冲的原始信号. 使用 Gauss 随机矩阵, 并通过 $[0, 512]$ 上的均匀采样, 应用 Half 阈值迭代算法和基于 AMP 改进的 Half 阈值迭代算法重建信号 s . 实验结果见图 1(a)~(d). 在图 1(a)~(d) 中, Original 显示的是原始信号图形, Half (240) 显示的是 Half 阈值迭代算法通过 240 个采样所获得的重建信号图形, 此时重建信号与原始信号的总误差 (总误差 = $\|s - \hat{s}\|_2$, \hat{s} 为 s 的估计值) 为 $1.9774\text{E}-07$, Half (239) 显示的是 Half 阈值迭代算法通过 239 个采样所获得的重建信号图形, 此时重建信号与原始信号的总误差为 6.6391. Halfamp (239) 显示的是基于 AMP 改进的 Half 阈值迭代算法通过 239 个采样所获得的重建信号图形, 此时重建信号与原始信号的总误差为 $6.7583\text{E}-08$. 从图 1(a)~(d) 可以看出: 采样数从 240 减少到 239 时, Half 阈值迭代算法已经不能完全重建原始信号, 然而基于 AMP 改进的 Half 阈值迭代算法却能完全重建.

接下来, 分别采样 235 和 233, 然后分别应用 Half 阈值迭代算法和基于 AMP 改进的 Half 阈值迭代算法重建信号 s . 实验结果如图 1(e)~(h) 所示, Half (235) 显示的是 Half 阈值迭代算法通过 235 个采样所获得的重建信号图形, 此时重建信号与原始信号的总误差为 11.2340. Halfamp (235) 显示的是基于 AMP 改进的 Half 阈值迭代算法通过 235 个采样所获得的重建信号图形, 此时重建信号与原始信号的总误差为 $9.5819\text{E}-08$. Half (233) 显示的是 Half 阈值迭代算法通过 233 个采样所获得的重建信号图形, 此时重建信号与原始信号的总误差为 9.3684. Halfamp (233) 显示的是基于 AMP 改进的 Half 阈值迭代算法通过 233 个采样所获得的重建信号图形, 此时重建信号与原始信号的总误差为 $6.7636\text{E}-08$.

从上述实验可以看出: 对于采样数小于 240, Half 阈值迭代算法均不能完全重建原始信号, 然而, 对于基于 AMP 改进的 Half 阈值迭代算法, 直到采样数减少到 233 仍能重建原始信号. 说明基于 AMP

表 1 Half 阈值迭代算法与基于 AMP 改进的 Half 阈值迭代算法信号重建的实验结果

Table 1 The signal reconstruction results of Half iterative thresholding algorithm and Half iterative thresholding algorithm based on AMP

	$k=60$	$k=80$	$k=100$	$k=110$	$k=130$	$k=150$
Half	160	190	221	229	240	272
Halfamp	148	180	213	217	233	265

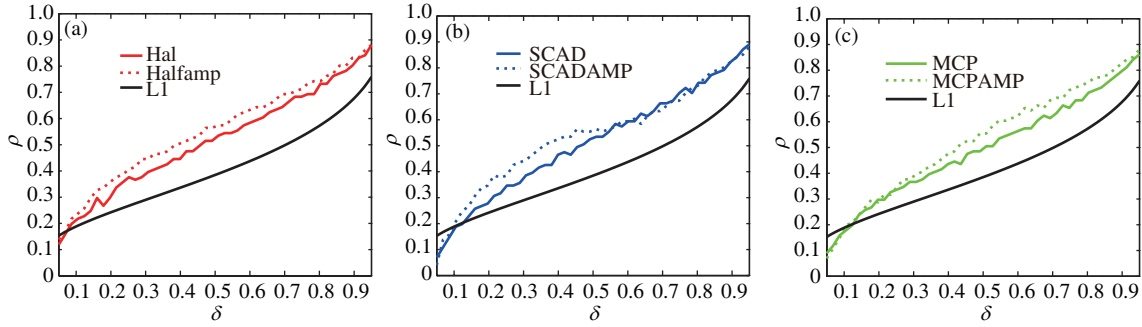


图 2 (网络版彩图) 三种非凸正则化方法的阈值迭代算法和其基于 AMP 改进的阈值迭代算法的相变实验结果图

Figure 2 (Color online) The phase transition results of the iterative thresholding algorithm and its improved version based on AMP for solving three nonconvex regularization problem. (a) Half iterative thresholding algorithm and Half iterative thresholding algorithm based on AMP; (b) SCAD iterative thresholding algorithm and SCAD iterative thresholding algorithm based on AMP; (c) MCP iterative thresholding algorithm and MCP iterative thresholding algorithm based on AMP. In each subfigure, we also showing the theoretical phase transition of the L_1 regularization

改进的 Half 阈值迭代算法的重建能力比 Half 阈值迭代算法的重建能力强。

为了进一步判断基于 AMP 改进的 Half 阈值迭代算法能否提高 Half 阈值迭代算法的重建能力, 我们进行了下面的实验: 分别取含有 60, 80, 100, 110, 130, 150 个非零窄脉冲的原始信号. 使用 Gauss 随机矩阵, 并通过 $[0, 512]$ 上的均匀采样, 使用 Half 阈值迭代算法和基于 AMP 改进的 Half 阈值迭代算法重建信号 s . 我们统计了不同信号稀疏度的两种方法重建信号需要的最少采样个数, 实验以 $\|s - \hat{s}\|_2 \leq 10^{-4}$ (\hat{s} 为 s 的估计值) 作为信号重建成功的判断标准. 实验结果见表 1, 实验结果显示, 基于 AMP 改进的 Half 阈值迭代算法提高了 Half 阈值迭代算法的重建能力。

实验 2 通过相变方法说明基于 AMP 改进的 Half 阈值迭代算法的稀疏能力. Donoho 在文献 [23] 中对于 L_0/L_1 等价性展开了研究, 也即研究 L_1 正则化的稀疏重建能力, 研究结果请参考文献 [23] 中的图 1.1: 图中的曲线是 L_1 正则化的相变曲线, 图中每一个点对应于向量 (n, k, p) (即信号的维数 p , 采样的个数 n , 以及信号的稀疏度 k), 因此, 每个点对应其相应的信号恢复问题. L_1 正则化能从理论上精确重建曲线下方的点所对应的原始信号, 而不能完全重建曲线上方的点所对应的原始信号, 进而从理论上提出了一种 L_1 正则化稀疏重建能力的度量方式. Zhang 等 [24] 通过构造相变平台比较了不同正则化的稀疏信号重建能力。

下面通过相变方法具体研究 Half 阈值迭代算法和基于 AMP 改进的 Half 阈值迭代算法的稀疏能力. 首先固定信号的维数 $N = 256$, 然后采样数 p 取值从 1 到 N , 对应于每个采样数, 信号的稀疏度 k 取值从 1 到 N , 对每一组 (p, N, k) , 采用上述不同方法重建稀疏信号, 对每一个信号问题, 重复实验 30 次, 并且统计其精确重建成功次数, 如果成功次数大于 15 次, 则该信号可精确重建, 对于每组 (p, N) , 统计其精确重建信号所需的最小稀疏度 k . 实验以 $\frac{\|s - \hat{s}\|_2}{\|\hat{s}\|_2} \leq 10^{-4}$ (\hat{s} 为 s 的估计值) 作为信号重建成功的判断标准. 实验结果如图 2(a) 所示, 实验结果显示基于 AMP 改进的 Half 阈值迭代算法提高了

Half 阈值迭代算法的稀疏重建能力.

在实验中, 假设已知信号的稀疏度为 k , 那么, 对于

$$z_1^{(t+1)} = (I - A^T A)s^{(t)} + A^T A s, \quad (3.3)$$

$$z_2^{(t+1)} = (I - A^T A)s^{(t)} + A^T A s + \frac{1}{\delta} A^T r^{(t-1)} \langle \eta'_{t-1}(A^T r^{(t-1)} + s^{(t-1)}) \rangle, \quad (3.4)$$

把 $z_1^{(t+1)}$ 和 $z_2^{(t+1)}$ 分别按绝对值降序排列记为 $C_1^{(t+1)}$ 和 $C_2^{(t+1)}$. 我们的 λ 选取情况为: Half 阈值迭代算法选取 $\lambda = (\frac{4}{3}C_1^{(t+1)}(k+1))^{\frac{2}{3}}$, 基于 AMP 改进的 Half 阈值迭代算法选取 $\lambda = (\frac{4}{3}C_2^{(t+1)}(k+1))^{\frac{2}{3}}$.

进一步, 我们还研究了基于 AMP 改进的 SCAD, MCP 阈值迭代算法的相变实验. 考虑到篇幅问题, 我们这里仅给出实验结果. 实验结果如图 2(b) 和 (c) 所示, 实验结果显示基于 AMP 改进的 SCAD, MCP 阈值迭代算法提高了 SCAD, MCP 阈值迭代算法的稀疏重建能力.

4 结论

本文研究基于 AMP 的 $L_{\frac{1}{2}}$ 正则化方法, 采用 Belief Propagation 算法的思想构造基于 AMP 改进的 Half 阈值迭代算法, 并证明了在向量稀疏度和测量矩阵的相关性满足一定条件的情况下, 基于 AMP 改进的 Half 阈值迭代算法至多需要有限步就能精确估计稀疏向量. 从而从理论上为基于 AMP 改进的 Half 阈值迭代算法提供了支撑. 进一步, 我们通过实验具体分析了 $L_{\frac{1}{2}}$ 正则化方法的基于 AMP 改进的 Half 阈值迭代算法, 实验表明, 基于 AMP 改进的 Half 阈值迭代算法与 Half 阈值迭代算法相比, 估计精度更高, 从而具有广泛的应用价值. 对于几种重要的非凸正则化方法 (SCAD, MCP), 我们通过相变实验初步验证了基于 AMP 改进的 SCAD, MCP 阈值迭代算法能有效地提高 SCAD, MCP 阈值迭代算法的稀疏重建能力.

参考文献

- 1 Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle//Selected Papers of Hirotugu Akaike. New York: Springer, 1998. 199–213
- 2 Schwarz G. Estimating the dimension of a model. Ann Stat, 1978, 6: 461–464
- 3 Mallows C L. Some comments on C_p . Technometrics, 1973, 15: 661–675
- 4 Tibshirani R. Regression shrinkage and selection via the lasso. J Royal Stat Soc Ser B, 1996, 58: 267–288
- 5 Donoho D L, Huo X. Uncertainty principles and ideal atomic decomposition. IEEE Trans Inform Theory, 2001, 47: 2845–2862
- 6 Donoho D L, Elad E. Maximal sparsity representation via L_1 minimization. Proc Nat Aca Sci, 2003. 100: 2197–2202
- 7 Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. SIAM Review, 2001, 43: 129–159
- 8 Candes E J, Tao T. The dantzig selector: Statistical estimation when p is much larger than n . Quality Contr Appl Stat, 2009, 54: 83–84
- 9 Fan J Q, Li R Z. Statistical challenges with high dimensionality: feature selection in knowledge discovery. arXiv:math/0602133
- 10 Zhang C H. Nearly unbiased variable selection under minimax concave penalty. Ann Stat, 2010, 38: 894–942
- 11 Xu Z B, Zhang H, Wang Y, et al. $L_{\frac{1}{2}}$ regularization. Sci China Inf Sci, 2010, 53: 1159–1169
- 12 Zhang H, Xu Z B, Wang Y, et al. A sharp nonasymptotic bound and phase diagram of $L_{\frac{1}{2}}$ regularization. Acta Math Sin, 2014, 30: 1242–1258
- 13 Liang Y, Liu C, Luan X Z, et al. Sparse logistic regression with a $L_{\frac{1}{2}}$ penalty for gene selection in cancer classification. BMC Bioinformatics, 2013, 14: 1–12

- 14 Xu Z B, Chang X Y, Xu F M, et al. $L_{\frac{1}{2}}$ regularization: A thresholding representation theory and a fast solver. *IEEE Trans Neural Netw Learn Syst*, 2012, 23: 1013–1027
- 15 Jordan M I. Graphical models. *Stat Sci*, 2004, 19: 140–155
- 16 Donoho D L, Maleki A, Montanari A. Message-passing algorithms for compressed sensing. *Proc Natl Acad Sci*, 2009, 106: 18914–18919
- 17 Donoho D L, Maleki A, Montanari A. How to design message passing algorithms for compressed sensing. Preprint, 2011
- 18 Candes E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Commun Pure Appl Math*, 2006, 59: 1207–1223
- 19 Donoho D L. Compressed sensing. *IEEE Trans Inf Theory*, 2006, 52: 1289–1306
- 20 Majumdar A, Ward R K. Compressed sensing of color images. *Signal Process*, 2010, 90: 3122–3127
- 21 Shi G M, Liu D H, Gao D H, et al. Advances in theory and application of compressed sensing. *Acta Electron Sin*, 2009, 37: 1070–1081 [石光明, 刘丹华, 高大化, 等. 压缩感知理论及其研究进展. *电子学报*, 2009, 37: 1070–1081]
- 22 Zhang H, Liang Y, Xu Z B, et al. Compressive sensing with noise based on SCAD penalty. *Acta Math Sin*, 2013, 56: 767–776
- 23 Donoho D L. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput Geom*, 2006, 35: 617–652
- 24 Zhang H, Liang Y, Gou H L, et al. The essential ability of sparse reconstruction of different compressive sensing strategies. *Sci China Inf Sci*, 2012, 55: 2582–2589

Approximate message passing algorithm for $L_{1/2}$ regularization

Hui ZHANG¹ & Hai ZHANG^{1,2*}

1. *School of Mathematics, Northwest University, Xi'an 710069, China;*

2. *Institute of Applied Mathematics, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing 100190, China*

*Corresponding author. E-mail: zhanghai@nwu.edu.cn

Abstract In this paper, we study the approximate message passing (AMP) algorithm for $L_{\frac{1}{2}}$ regularization. We propose an improved half iterative thresholding algorithm, which is inspired by belief propagation in graphical model. Further, we study the convergence property of the new algorithm. Through extensive simulations, we show that the iterative thresholding algorithms based on AMP for several nonconvex regularization approaches have strong reconstruction capabilities and high phase transition.

Keywords $L_{\frac{1}{2}}$ regularization, half iterative thresholding algorithm, AMP algorithm, variable selection, sparsity



Hui ZHANG was born in 1990. She is currently working for an M.S. degree in Statistics from Northwest University, Xi'an, China. Her current research interests include high-dimensional statistics and compressive sensing.



Hai ZHANG was born in 1975. He obtained his Ph.D. degree in Applied Mathematics from Xi'an Jiaotong University, Xi'an, in 2012. He is currently a professor in the Department of Financial Mathematics and Statistics at Northwest University. His research interests include statistical machine learning, high-dimensional statistics, and social network analysis.