



基于稀疏表示的球面梯度下降算法

赵永红¹, 沈 益¹, 胡瑞芳²

(1. 浙江理工大学理学院, 杭州 310018; 2. 嘉兴学院南湖学院数理系, 浙江 嘉兴 314000)

摘 要: 针对基于稀疏表示的球面最小化问题, 结合梯度下降、球面投影、稀疏逼近等方法设计了球面上的迭代硬阈值(Iterative hard thresholding, IHT)算法。首先证明了该算法产生的序列收敛到模型的 L 稳定点, 然后通过 Nesterov 加速进一步提升了该算法的性能, 最后将加速后的算法应用于稀疏主成分分析(Sparse principal component analysis, SPCA)和 1-Bit 压缩感知(1-Bit compressive sensing, 1-Bit CS)。采用高斯随机矩阵进行测试, 并与 1-Bit CS 中的二进制迭代硬阈值算法、SPCA 算法中的截断幂法进行了对比, 数值实验表明: 该算法可以有效地求解基于稀疏表示的球面最小化问题, 算法产生的序列收敛到优化模型的 L 稳定点, 加速后算法的收敛速度优于原求解算法。

关键词: 稀疏约束; 球面 IHT; 1-Bit 压缩感知; 稀疏主成分分析

中图分类号: TP391.41

文献标志码: A

文章编号: 1673-3851(2020)05-0714-08

Spherical gradient descent method based on sparse representation

ZHAO Yonghong¹, SHEN Yi¹, HU Ruifang²

(1. School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. Department of Mathematics and Information Engineering, Nanhu College, Jiaxing University, Jiaxing 314001, China)

Abstract: To solve the spherical minimization problem based on sparse representation, the iterative hard thresholding (IHT) algorithm on the sphere was proposed by combining gradient decent, projection on the sphere and sparse approximation. First, the sequence convergence to the L stationary points of the model was proven. Then, the Nesterov acceleration was used to further improve the performance of this algorithm. Finally, the accelerated algorithm was applied to sparse principal component analysis (SPCA) and 1-Bit Compressive Sensing (1-Bit CS). Gaussian random matrix was used to test the performance of the improved algorithm, and the proposed algorithm was compared with the binary iterative hard thresholding algorithm in 1-Bit CS and truncated power method in SPCA. Numerical experiments show that the proposed algorithm can effectively solve the spherical minimization problem based on sparse representation; the sequence generated by the algorithm converges to the L stationary points of the optimization model, and the convergence speed of the accelerated algorithm is faster than that of the original algorithm.

Key words: sparsity constraints; spherical IHT; 1-Bit compressive sensing; sparse principal component analysis

0 引 言

近年来, 基于稀疏表示以及稀疏重构的各类算法已成为优化领域的热点之一, 在信号处理、计算机视觉、推荐系统、生物信息学等领域具有广泛的应用^[1]。稀疏约束项的引入给模型的求解、算法的分析带来了挑

收稿日期: 2020-01-09 网络出版日期: 2020-06-03

基金项目: 浙江省自然科学基金杰出青年基金项目(LR19A010001); 嘉兴学院南湖学院重点研究项目(N41472001-40)

作者简介: 赵永红(1994—), 女, 安徽淮北人, 硕士研究生, 主要从事小波分析与压缩感知方面的研究。

通信作者: 胡瑞芳, E-mail: ruifanghu@qq.com

战^[2-3]。针对带有稀疏约束项的主成分分析问题,Zou 等^[2]提出了 SPCA 数学模型。对于给定的数据集,SPCA 问题可建模为:

$$\begin{cases} \min_{x \in \mathbb{R}^n} -x^T A^T A x, \\ \text{s.t. } \|x\| = 1, \quad \|x\|_0 \leq s. \end{cases}$$

对于 SPCA 模型,Luss 等^[4-5]采用了凸松弛的方法进行近似求解;Yuan 等^[6]提出了截断幂法,并在高斯噪声的作用下算法具有较好的收敛性;Merola 等^[7]将数据矩阵的主成分投影到变量的子集上,提出了新模型以代替 SPCA 模型;Hager 等^[8]采用了梯度投影法和近似牛顿法求解 SPCA 模型,给出了算法收敛的充分条件。

压缩感知理论是一种基于信号的稀疏性的采样理论。针对测量值仅保留符号信息,丢失了信号幅度信息的情况,Boufounos 等^[3]提出了 1-Bit CS 理论。Jacques 等^[9]采用了 BIHT 算法恢复稀疏信号,他们认为,已有的数学理论无法分析 BIHT 算法的收敛性。许多学者对 BIHT 算法进行了改进和推广,如:Yan 等^[10]提出了可以识别采样值符号变换的自适应异常值追踪算法,该算法具有很强的自适应性;Kamilov 等^[11]提出了自适应阈值方法;Shen 等^[12]通过光滑函数逼近的方式,提高了 BIHT 算法的性能;Liu 等^[13]提出了基于稀疏约束的次梯度下降算法,刻画了单侧 1 范数的 BIHT 算法的收敛性。尽管 BIHT 算法已有许多研究成果,然而对于单侧 2 范数的 BIHT 算法仍然缺少理论上的收敛性分析。

本文研究基于稀疏表示的球面最小化问题的数学模型为:

$$\begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } \|x\| = 1, \quad \|x\|_0 \leq s \end{cases} \quad (1)$$

其中: $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是连续可微的函数; s 是大于零小于 n 的正整数; $\|\cdot\|$ 为欧氏空间的 l_2 范数, $\|x\|_0$ 表示向量 x 中非零元素的个数。首先,本文提出了球面上的迭代硬阈值算法来求解模型(1)。该算法是一种梯度投影方法,其中硬阈值投影和球面投影两个步骤可以转化为非线性稀疏逼近的问题。本文用优化理论和迭代硬阈值算法的分析方法^[14-15],对提出的球面上的迭代硬阈值算法给出了收敛性分析。然后,将该结论应用于 1-Bit CS 和 SPCA 的几类算法,证明了单侧 2 范数 BIHT 算法和球面梯度下降算法产生的序列均收敛到对应模型的 L 稳定点。最后,通过 Nesterov 加速^[16]提升了算法的准确率和收敛速度。

1 球面上的稀疏约束问题的基本知识

受 Beck 等^[16]工作的启发,本文考虑优化问题的形式如下:

$$\min\{f(x): x \in \Omega\},$$

其中: $\Omega = \{x \in \mathbb{R}^n: \|x\| = 1, \|x\|_0 \leq s\}$ 表示约束在球面上最多有 s 个非零元素的向量全体。给定集合 Ω 和向量 x , 记 x 在 Ω 上的投影为:

$$P_\Omega(x) = \arg \min_{y \in \Omega} \|y - x\| \quad (2)$$

命题 1^{[17]Proposition 2.1} 给定集合 Ω , 可得到

$$\frac{T_s(x)}{\|T_s(x)\|} \in P_\Omega(x) \quad (3)$$

其中: $T_s(x)$ 表示保留向量 x 中绝对值最大的 s 个分量,其余分量替换成 0。

注 1 Ω 是非凸集合,投影 $P_\Omega(\cdot)$ 不是单值的。

定义 1 对任意向量 $x^* \in \Omega$, 如果满足

$$x^* \in P_\Omega\left(x^* - \frac{1}{L} \nabla f(x^*)\right),$$

则称 x^* 是模型(1)的 L 稳定点。

假设 1 $f(x)$ 是二阶连续可微的且其梯度 $\nabla f(x)$ 是 Lipschitz 连续的,即:

$$\|\nabla f(x) - \nabla f(y)\| \leq L(f) \|x - y\|, \forall x, y \in \Omega.$$

其中 $L(f)$ 是 Lipschitz 常数。

引理 1^{[8] lemma 2.2} 如果 $x_k \in \Omega$, 对于所有的 $y \in P_\Omega\left(x_k - \frac{1}{L} \nabla f(x_k)\right)$, 则

$$\nabla f(x_k)^T(y - x_k) \leq 0.$$

当 $y = x_{k+1}$ 时, 对于所有的 $x_k \in P_\Omega\left(x_k - \frac{1}{L} \nabla f(x_k)\right)$, 得出

$$\nabla f(x_k)^T(x_{k+1} - x_k) \leq 0.$$

引理 2^{[17] lemma 2.1} 假设 1 成立, 对于任意 $L \geq L(f)$, 有 $f(y) \leq h_L(y, x)$, 其中

$$h_L(y, x) \equiv f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2.$$

引理 3 假设 1 成立, 对于任意的 $L > L(f)$, 则对于任意 $x, y \in \Omega$ 满足式

$$y \in P_\Omega\left(x - \frac{1}{L} \nabla f(x)\right),$$

得到:

$$f(x) - f(y) \geq \frac{L - L(f)}{2} \|x - y\|^2.$$

证明 由引理 2 知道

$$\begin{aligned} h_L(z, x) &= f(x) + \langle \nabla f(x), z - x \rangle + \frac{L}{2} \|x - z\|^2 = \\ &= \frac{L}{2} \|z - (x - \frac{1}{L} \nabla f(x))\|^2 + f(x) - \frac{1}{2L} \|\nabla f(x)\|^2 \end{aligned} \quad (4)$$

这里 $f(x) - \frac{1}{2L} \|\nabla f(x)\|^2$ 关于 z 为常数。

又因为 $y \in P_\Omega\left(x - \frac{1}{L} \nabla f(x)\right)$, 所以

$$y \in \arg \min_{z \in \Omega} \left\| z - \left(x - \frac{1}{L} \nabla f(x)\right) \right\|^2,$$

故式(4)等价于 $y \in \arg \min_{z \in \Omega} h_L(z, x)$, 即得到 $h_L(y, x) \leq h_L(x, x) = f(x)$ 。

通过引理 2, 得到

$$f(x) - f(y) \geq f(x) - h_{L(f)}(y, x) \geq h_L(y, x) - h_{L(f)}(y, x).$$

通过定义知道

$$h_{L(f)}(y, x) = h_L(y, x) - \frac{L - L(f)}{2} \|x - y\|^2,$$

则

$$f(x) - f(y) \geq \frac{L - L(f)}{2} \|x - y\|^2.$$

引理 4 如果假设 1 成立, $L > L(f)$, 若 x^* 为模型(1)的最优解, 满足

$$x^* = P_\Omega\left(x^* - \frac{1}{L} \nabla f(x^*)\right),$$

则称 x^* 是 L 稳定点。

证明 假设存在向量 $y \in P_\Omega\left(x^* - \frac{1}{L} \nabla f(x^*)\right)$ 并且 $y \neq x^*$ 。通过引理 3 得到

$$f(x^*) - f(y) \geq \frac{L - L(f)}{2} \|x^* - y\|^2.$$

这里缩小了 x^* 的最优性。则 x^* 是集合 $P_\Omega\left(x^* - \frac{1}{L} \nabla f(x^*)\right)$ 中唯一的向量且为 L 稳定点。

2 球面上的迭代硬阈值算法及收敛性分析

下面给出本文提出的球面上的迭代硬阈值算法, 即算法 1。

输入: Lipschitz 常数 $L \geq L(f)$

初始化: $\mathbf{x}_0 \in \Omega$, 令 $k=0$

while 不收敛 do

$$\mathbf{x}_{k+1} \in P_{\Omega}\left(\mathbf{x}_k - \frac{1}{L} \nabla f(\mathbf{x}_k)\right)$$

$$k = k + 1$$

end while

输出: \mathbf{x}_{k+1}

引理 5 $f(\mathbf{x})$ 是满足假设 1 的连续可微的函数, 如果 $\{\mathbf{x}_k\}$ 是算法 1 产生的序列, 则

$$\text{a) } f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{L - L(f)}{2} \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2;$$

b) $\{f(\mathbf{x}_k)\}$ 是非增序列;

$$\text{c) } \|\mathbf{x}_k - \mathbf{x}_{k+1}\| \rightarrow 0;$$

d) 对于每个 $k=0, 1, 2, \dots$, 如果 $\mathbf{x}_k \neq \mathbf{x}_{k+1}$, 则 $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$;

e) 如果序列 \mathbf{x}_k 的极限是 \mathbf{x}^* , 对于所有的 $\mathbf{y} \in P_{\Omega}\left(\mathbf{x}^k - \frac{1}{L} \nabla f(\mathbf{x}^k)\right)$, 则

$$\nabla f(\mathbf{x}^k)^T (\mathbf{y} - \mathbf{x}^k) \leq 0.$$

证明 引理 3 证明了对于任意的 $\mathbf{x} \in \Omega, \mathbf{y} \in P_{\Omega}\left(\mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x})\right)$ 时

$$f(\mathbf{x}) - f(\mathbf{y}) \geq \frac{L - L(f)}{2} \|\mathbf{x} - \mathbf{y}\|^2$$

成立。因此取 $\mathbf{x} = \mathbf{x}_k, \mathbf{y} = \mathbf{x}_{k+1}$,

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{L - L(f)}{2} \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 \geq 0 \quad (5)$$

成立, 即

$$f(\mathbf{x}_k) \geq f(\mathbf{x}_{k+1}) \geq 0.$$

综上所述, 函数列单调下降且有下界因此收敛, 即

$$\lim_{k \rightarrow \infty} [f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})] = 0,$$

结合式(5)的性质知 $\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \rightarrow 0$ 成立。引理的其余性质从式(5)直接可证。

定理 1 令 $f(\mathbf{x})$ 为满足假设 1 的连续可微的函数, $\{\mathbf{x}_k\}$ 是算法 1 产生的迭代序列, 则序列 $\{f(\mathbf{x}_k)\}_{k \geq 0}$ 收敛。

证明 由引理 5 可知道 $\{f(\mathbf{x}_k)\}_{k \geq 0}$ 是单调且有下界的, 故 $\{f(\mathbf{x}_k)\}_{k \geq 0}$ 收敛。

定理 2 $f(\mathbf{x})$ 是满足假设 1 的连续可微的函数, 如果 $\{\mathbf{x}_k\}$ 是算法 1 产生的迭代序列, 则序列 $\{\mathbf{x}_k\}_{k \geq 0}$ 任意一个聚点都是 L 稳定点。

证明 假设 \mathbf{x}^* 是序列的聚点, 则存在子序列 $\{\mathbf{x}_{k_m}\}_{m \geq 0}$ 收敛到 \mathbf{x}^* 。通过引理 3 得到:

$$f(\mathbf{x}_{k_m}) - f(\mathbf{x}_{k_m+1}) \geq \frac{L - L(f)}{2} \|\mathbf{x}_{k_m} - \mathbf{x}_{k_m+1}\|^2.$$

由于 $\{f(\mathbf{x}_{k_m})\}_{m \geq 0}$ 和 $\{f(\mathbf{x}_{k_m+1})\}_{m \geq 0}$ 都收敛到 f^* , 则

$$\lim_{m \rightarrow \infty} [f(\mathbf{x}_{k_m}) - f(\mathbf{x}_{k_m+1})] = 0,$$

结合上述不等式得到 $\lim_{m \rightarrow \infty} \mathbf{x}_{k_m+1} = \mathbf{x}^*$ 。对于所有的 $n \geq 0$

$$\mathbf{x}_{k_m+1} \in P_{\Omega}\left(\mathbf{x}_{k_m} - \frac{1}{L} \nabla f(\mathbf{x}_{k_m})\right),$$

令 $\mathbf{z}^* = \mathbf{x}_{k_m} - \frac{1}{L} \nabla f(\mathbf{x}_{k_m})$, 则

$$\mathbf{x}_{k_{m+1}} = \frac{T_s(\mathbf{z}^*)}{\|T_s(\mathbf{z}^*)\|}.$$

通过连续算子 $P_\Omega(\cdot)$, 取 $m \rightarrow \infty$, 得到

$$\mathbf{x}^* \in P_\Omega\left(\mathbf{x}^* - \frac{1}{L} \nabla f(\mathbf{x}^*)\right).$$

因此聚点是 L 稳定点。

注2 单侧2范数的BIHT算法^[9]和Hager等^[8]采用梯度投影法均为算法1的特例,因此定理2的结果说明,两类算法产生的序列均收敛到对应模型的 L 稳定点。

在算法1的基础上,借鉴Accelerated proximal gradient method^[16],通过Nesterov加速提升算法1的收敛速度。为了验证加速带来的优势,本文以1-Bit CS和SPCA为例比较不同算法的性能。算法2(算法SIHT)的具体描述如下:

输入: Lipschitz 常数 $L \geq L(f)$, $t_1 = 1, k = 1, y_1 = x_1$

while 不收敛 do

$$\mathbf{x}_{k+1} \in P_\Omega\left(\mathbf{y}_k - \frac{1}{L} \nabla f(\mathbf{y}_k)\right)$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$\mathbf{y}_{k+1} = \mathbf{x}_{k+1} + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

$$k = k + 1$$

end while

$$\text{输出: } \hat{\mathbf{x}} = \frac{\mathbf{y}_n}{\|\mathbf{y}_n\|}$$

3 实验结果与分析

本文通过数值实验验证球面迭代硬阈值算法的有效性。在一台安装MATLAB R2016a的电脑上进行数值实验。操作系统为Windows10 1909版本,主要配置为Core i3 9100F CUP (3.7 GHz)和8 GB内存。

3.1 1-Bit CS

压缩感知理论假设信号具有稀疏性,在已知测量矩阵和线性测量信息的情况下,通过求解凸优化模型恢复稀疏信号^[18-19]。实际问题中的测量值精度是有限的,需要在测量值量化的假设下建立恢复稀疏信号的理论。因此,Boufounos等^[3]提出1-Bit CS问题:

$$\mathbf{y} = \text{sgn}(\mathbf{z}) = \text{sgn}(\Phi\mathbf{x}) \quad (6)$$

其中: $\text{sgn}(\cdot)$ 表示符号函数。

非线性系统(6)是病态的,无法直接求解。Jacques等^[9]提出模型:

$$\begin{cases} \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} \odot \Phi\mathbf{x}\|_2^2 \\ \text{s.t. } \|\mathbf{x}\|_2 = 1, \quad \|\mathbf{x}\|_0 \leq s \end{cases} \quad (7)$$

其中:任意的两个向量 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} \odot \mathbf{y}$ 表示 Hadamard 乘积:

$$(\mathbf{x} \odot \mathbf{y})_i = x_i y_i, \quad i = 1, 2, \dots, n.$$

负函数 $[\cdot]_-$ 的第 i 项表示

$$([\mathbf{x}]_-)_i = -\min(x_i, 0) = \begin{cases} -x_i, & x_i < 0 \\ 0, & x_i \geq 0 \end{cases}.$$

模型(7)通过迭代算法

$$a_{k+1} = x_k - (Y\Phi)^T \cdot [Y\Phi x_k]_{-},$$

$$x_{k+1} = T_s(a_{k+1})$$

求解,其中 $Y = \text{diag}(y)$ 为测量向量 y 构成的对角矩阵。算法被称为单侧 2 范数的 BIHT 算法^[8]。

为比较算法的性能,试验中对高斯随机矩阵的规模设定为 m 分别取 500、1000, n 分别取 800、1000。对于给定的随机矩阵 Φ , 常数取值 L 为矩阵 $\Phi^T \Phi$ 的最大特征值。信号的稀疏度 s 就是信号中非零元素的个数, s 分别取 10、15。 x 为原始信号, 算法重建得到 x' , 重建误差取为:

$$\left\| \frac{x}{\|x\|} - \frac{x'}{\|x'\|} \right\|.$$

$n=800, m=500$ 时不同稀疏度下重建误差曲线如图 1 所示。由图 1(a)可知,当 $s=10$ 时,本文的 SIHT 算法的收敛速度是单侧 2 范数的 BIHT 算法的 3 倍左右,最终的重构误差稳定在 0.05 左右,但在相同的稀疏度下,由于每次实验产生不同随机矩阵导致最终呈现的不同重构误差;由图 1(b)可知,当 $s=15$ 时,SIHT 算法的收敛速度大概是单侧 2 范数的 BIHT 算法的 3 倍,最终的重构误差稳定在 0.08 左右。

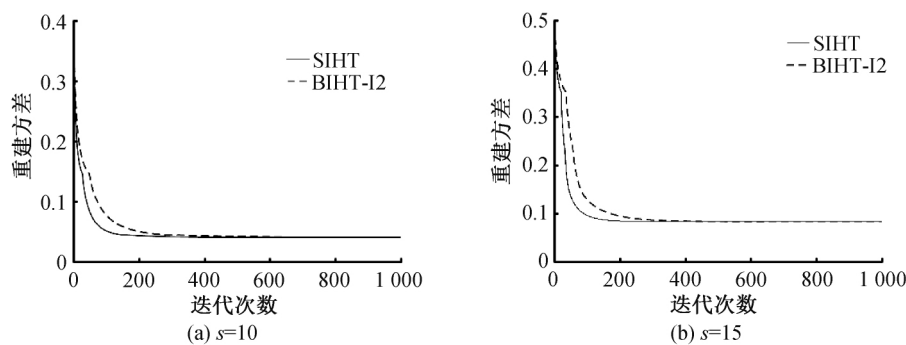


图1 $n=800, m=500$ 时不同稀疏度下信号重建的误差曲线

$n=1000, m=1000$ 时,不同稀疏度下重建误差曲线如图 2 所示。由图 2(c)可知, $s=10$ 时,SIHT 算法的收敛速度是单侧 2 范数的 BIHT 算法的 3 倍左右,算法达到收敛且收敛结果比较稳定,最终的重构误差稳定在 0.04 左右;由图 2(d)知, $s=15$ 时,SIHT 算法的收敛速度大概是单侧 2 范数的 BIHT 算法的 2 倍左右,算法达到收敛且收敛结果比较稳定,并且重构误差大约稳定在 0.075 左右。

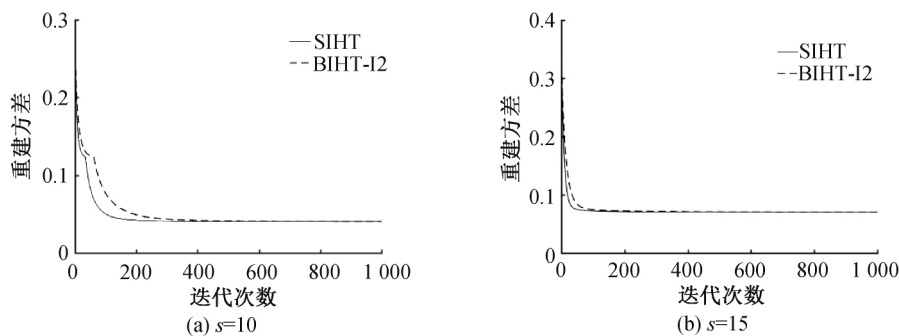
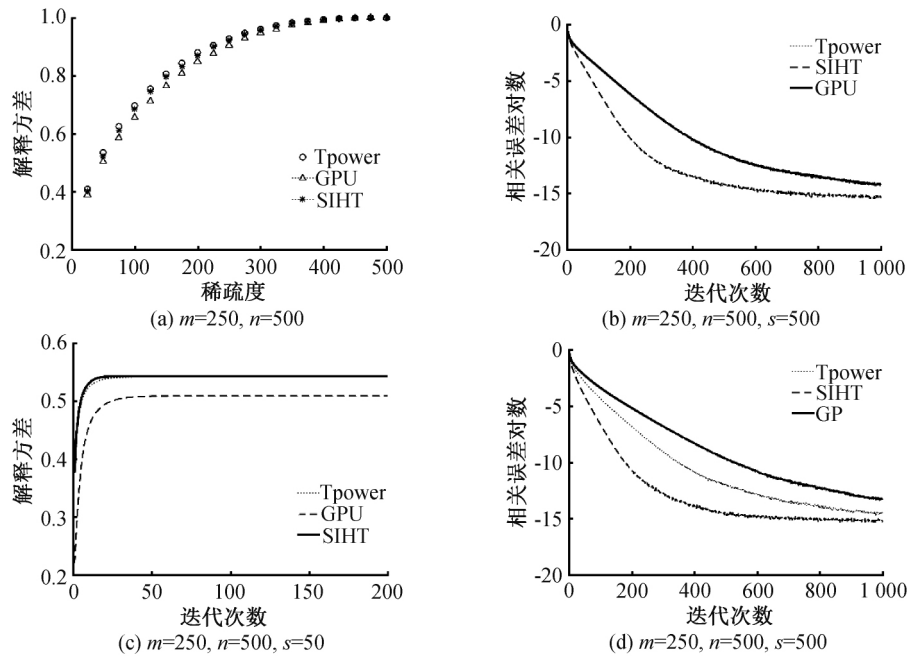


图2 $s=1000, m=1000$ 时不同稀疏度下信号重建的误差曲线

3.2 SPCA

本文将球面上的迭代硬阈值法的性能与截断幂法 (Truncated power method, Tpower), 具有单位梯度的投影算法的性能进行比较。本节的数值实验基于 SPCA 问题, 用比率 $\frac{v^T \Sigma v}{y^T \Sigma y}$ 来比较算法的性能, v 是算法求解出来的稀疏的第一主成分, y 是第一主成分。这个比率成为解释方差比。实验考虑随机产生的数据, 其中 $\Sigma = A^T A$, 数据矩阵 $A \in \mathbb{R}^{m \times n}$ 是高斯矩阵。类似地, 常数 L 取为 $A^T A$ 的最大特征值。本文中呈现的每个结果均基于 100 个随机矩阵的平均值。不同 m 和 s 变化时的误差曲线如图 3 所示。

由图 3(a) 知, 当 $m=250, n=500$ 时, 随着稀疏度的降低单位梯度投影算法 (Gradient projection

图3 不同 m 和 s 变化时的误差曲线

algorithm with unit stepsize, GPU) 和 Tpower 性能更好。当稀疏度接近 500 时, 所有算法产生相同的结果, 但算法的收敛性大不相同。为说明这点, 下面考虑 s 为 500 时, 其中矩阵 Σ 的最大特征值 λ_{\max} 对应的归一化的特征向量。对于任意算法产生的最优解的近似值 λ_{opt} 相对误差为:

$$\frac{|\lambda_{\max} - \lambda_{\text{opt}}|}{|\lambda_{\max}|}.$$

相关误差对数为:

$$\lg \frac{|\lambda_{\max} - \lambda_{\text{opt}}|}{|\lambda_{\max}|}.$$

由图 3(b) 知, SIHT 算法比 GPU 和 Tpower 具有更高的求解精度, 在 600 次迭代后 SIHT 算法的相对误差约为 10^{-15} , GPU 和 Tpower 相对误差约为 10^{-12} 。由图 3(c) 知, 当 $m=250, n=500, s=50$ 时, SIHT 算法比 Tpower 算法和 GPU 算法具有更好的收敛效果。由图 3(d) 知, 当 $m=250, n=500, s=500$ 时, SIHT 算法比 Tpower 算法和步长为 $1/L$ 梯度投影算法 (Gradient projection method, GP) 时具有更快的收敛速度和更高的求解精度。

4 结 论

本文研究了基于稀疏表示的球面最小化问题, 首先, 提出了球面上的迭代硬阈值算法, 证明了算法产生的序列收敛到模型的 L 稳定点; 然后, 通过 Nesterov 加速提高该算法的收敛速度; 最后, 采用高斯随机矩阵对 1-Bit 压缩感知和稀疏主成分分析进行实验。实验表明, 该算法收敛到优化模型的 L 稳定点, 并且加速后算法的收敛速度优于原来的优化方法, 是可以作为解决基于稀疏表示的球面最小化问题的有效方法。目前的算法无法处理复杂流形上的优化问题, 后续将研究一般流形上的稀疏优化问题。

参考文献:

- [1] Elad M. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing [M]. Springer, 2010: 169-181.
- [2] Zou H, Hastie T, Tibshirani R. Sparse principal component analysis [J]. Journal of Computational and Graphical Statistics, 2006, 15(2): 265-286.
- [3] Boufounos P T, Baraniuk R G. 1-Bit compressive sensing [C] // 2008 42nd Annual Conference on Information Sciences and

- Systems(CISS). Princeton, NJ, USA: IEEE, 2008, 3: 16-21.
- [4] Luss R, Teboulle M. Convex approximations to sparse PCA via lagrangian duality[J]. Operations Research Letters, 2011, 39(1):57-61.
- [5] Luss R, Teboulle M. Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint[J]. SIAM Review, 2013, 55(1): 65-98.
- [6] Yuan X T, Zhang T. Truncated power method for sparse eigenvalue problems[J]. Journal of Machine Learning Research, 2013, 14: 899 - 925
- [7] Merola G M, Chen G M. Projection sparse principal component analysis: An efficient least squares method[J]. Journal of Multivariate Analysis, 2019, 173: 366-382.
- [8] Hager W W, Phan D T, Zhu J J. Projection algorithms for nonconvex minimization with application to sparse principal component analysis[J]. Journal of Global Optimization, 2016, 65(4): 657-676.
- [9] Jacques L, Laska J N, Boufounos P T, et al. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors[J]. IEEE Transactions on Information Theory, 2013, 59(4): 2082-2102.
- [10] Yan M, Yang Y, Osher S. Robust 1-bit compressive sensing using adaptive outlier pursuit[J]. IEEE Transactions on Signal Processing, 2012, 60(7): 3868-3875.
- [11] Kamilov U S, Bourquard A, Amini A, et al. One-bit measurements with adaptive thresholds [J]. IEEE Signal Processing Letters, 2012, 19(10): 607-610.
- [12] Shen L X, Suter B W. one-bit compressive sampling via l_0 minimization[J/OL]. EURASIP Journal on Advances in Signal Processing, 2016, 71 (2016-06-14) [2020-01-09]. <https://doi.org/10.1186/s13634-016-0369-4>.
- [13] Liu D K, Li S, Shen Y. One-bit compressive sensing with projected subgradient method under sparsity constraints[J]. IEEE Transactions on Information Theory, 2019, 65(10):6650-6663.
- [14] Nesterov Y. Lectures on convex optimization[M]. Cham: Springer International Publishing, 2018:51-86.
- [15] Beck A, Eldar Y C. Sparsity constrained nonlinear optimization: Optimality conditions and algorithms[J]. SIAM Journal on Optimization, 2013, 23(3): 1480-1509.
- [16] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1):183-202.
- [17] Beck A, Hallak N. On the minimization over sparse symmetric sets: Projections, optimality conditions, and algorithms [J]. Mathematics of Operations Research, 2016, 41(1):196-223.
- [18] Candes E J, Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory, 2005, 51(12):4203-4215.
- [19] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4):1289-1306.

(责任编辑:康 锋)