

基于增广拉格朗日交替方向法的矩阵秩最小化算法研究

陈勇勇, 王永丽, 于慧慧

(山东科技大学 数学与系统科学学院, 山东 青岛 266590)

摘要:针对含有较大奇异值的矩阵秩最小化问题,采用对数行列式函数代替核范数作为秩函数的非凸近似,应用增广拉格朗日交替方向法求解矩阵秩最小化问题。当罚参数 $\beta > 1$ 时,证明此算法产生的迭代序列收敛到原问题的稳定点。最后利用实际数据和随机数据,通过数值实验验证所提出的算法较现有的求解核范数矩阵秩最小化问题的算法更高效。

关键词:对数行列式函数;核范数;增广拉格朗日交替方向法;低秩矩阵表示

中图分类号: N949

文献标志码: A

文章编号: 1672-3767(2016)04-0106-08

Matrix Rank Minimization Algorithm Based on Augmented Lagrangian Alternating Direction Method

CHEN Yongyong, WANG Yongli, YU Huihui

(College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

Abstract: To solve the matrix rank minimization problem with large singular values, the log-determinant function was used as a rank approximation instead of the nuclear norm and an augmented Lagrangian alternating direction method was proposed. When penalty parameter $\beta > 1$, the sequence of iterations generated by the proposed algorithm was proved to be convergent to a stationary point of the original problem. Finally, numerical experiments were conducted based on real data and random data. The results demonstrate that the proposed algorithm is more efficient than the existing nuclear norm method in solving the problem of matrix rank minimization.

Key words: log-determinant function; nuclear norm; augmented Lagrangian alternating direction method; low rank representation

1 引言

考虑矩阵秩最小化问题

$$\min_X \{rank(\mathbf{X}) : \mathbf{A}\mathbf{X} = \mathbf{B}\}. \quad (1)$$

其中, $\mathbf{X} \in \mathbf{R}^{m \times n}$, $\mathbf{A} \in \mathbf{R}^{p \times m}$, $\mathbf{B} \in \mathbf{R}^{p \times n}$ 是数据矩阵。

矩阵秩最小化问题在机器学习、数据挖掘、计算机视觉、图像处理、控制、生物信息学等领域有广泛的应

收稿日期: 2016-01-08

基金项目: 国家自然科学基金项目(11241005)

作者简介: 陈勇勇(1989—), 男, 山东泰安人, 硕士研究生, 研究方向为系统优化及应用、低秩矩阵恢复、分布式优化算法等。

王永丽(1977—)女, 山东栖霞人, 副教授, 博士, 研究方向为非线性优化理论与算法、分布式优化算法及数据挖掘等。

本文通信作者. E-mail: wangyongli@sdkt.net.cn

用,由于求解此问题是 NP-hard 问题,学者们往往使用核范数作为矩阵秩函数的凸近似^[15],即通过求解如下核范数最小化问题来得到问题(1)的近似解:

$$\min_X \{ \|X\|_* : AX = B \}. \quad (2)$$

其中, $\|g\|_*$ 表示矩阵的核范数,即矩阵的所有非零奇异值之和。但是在实际应用中,数据矩阵 B 通常都是带噪声的,因此求解问题(2)的一种常用的方法是最小二乘近似,即将问题(2)转化为如下核范数无约束最小化问题进行求解:

$$\min_X \left\{ \frac{1}{2\mu} \|AX - B\|_F^2 + \|X\|_* \right\}, \quad (3)$$

其中, $\mu > 0$ 为罚参数, $\|g\|_F$ 表示矩阵的 Frobenius 范数。

求解核范数无约束最小化问题常用的算法有加速近似梯度法 (accelerated proximal gradient method with line-search, APGL)^[1]、增广拉格朗日法 (augmented lagrangian method, ALM)^[2]、交替方向乘子法 (alternating direction method of multiplier, ADMM)^[3,4]、坐标下降法 (coordinate descent)^[5] 等。当矩阵的某些奇异值较大时,彭冲等^[7]求解子空间聚类问题时发现,利用对数行列式函数对矩阵秩函数进行近似的效果较好,而核范数的近似效果较差。基于此,本文用对数行列式函数代替核范数作为矩阵秩函数的近似,同时考虑到对数行列式函数的非凸性,应用增广拉格朗日交替方向法求解问题(3),提出一个新的算法,在适当条件下证明了算法的收敛性,并通过实例验证了算法的有效性。

2 对数行列式函数最小化问题

2.1 问题描述

在问题(3)中,利用对数行列式函数代替核范数,可得如下形式:

$$\min_X \left\{ \frac{1}{2\mu} \|AX - B\|_F^2 + G(X) \right\}, \quad (4)$$

其中: $G(X) := \log \det(I + X^T X) = \sum_i \log(1 + \sigma_i^2)$ 为对数行列式函数, σ_i 是矩阵 X 的奇异值。图 1 给出了核范数与对数行列式函数对“magic”矩阵秩的近似效果比较。

2.2 增广拉格朗日交替方向法

引入辅助变量 $Y \in \mathbb{R}^{m \times n}$, 将问题(4)转化成如下等价形式:

$$\min_{X, Y} \left\{ \frac{1}{2\mu} \|AY - B\|_F^2 + \log \det(I + X^T X) \right\} \quad (5)$$

$$s. t. \quad X - Y = 0$$

问题(5)的增广拉格朗日函数为

$$L_\beta(X, Y; \Lambda) = \frac{1}{2\mu} \|AY - B\|_F^2 +$$

$$\log \det(I + X^T X) - [\Lambda, X - Y] + \frac{\beta}{2} \|X - Y\|_F^2 \quad (6)$$

其中, $\beta > 0$ 是罚参数, Λ 是等式约束对应的拉格朗日乘子, I 是单位矩阵, $[\Lambda, B] = \text{trace}(\Lambda^T B)$ 表示标准的矩阵内积。给定任意初始矩阵 X_0 、 Λ_0 , 增广拉格朗日交替方向法的迭代框架为

$$Y_{k+1} = \arg \min_Y \{ L_\beta(X_k, Y; \Lambda_k) \}; \quad (7a)$$

$$X_{k+1} = \arg \min_X \{ L_\beta(X, Y_{k+1}; \Lambda_k) \}; \quad (7b)$$

$$\Lambda_{k+1} = \Lambda_k - \beta(X_{k+1} - Y_{k+1}). \quad (7c)$$

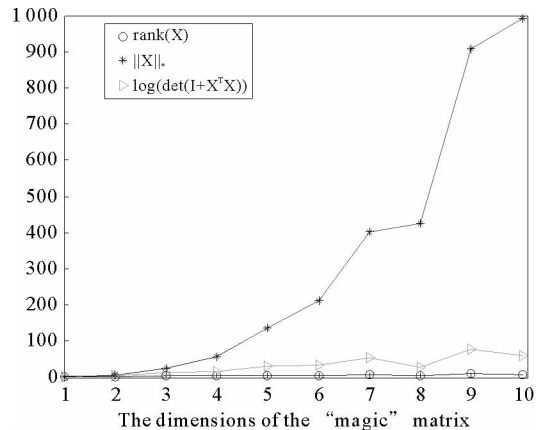


图 1 核范数与对数行列式函数对“magic”矩阵秩的近似效果比较

Fig. 1 The comparison of the behavior of the nuclear norm and the log-determinant function in approximating the rank of the “magic” matrix.

子问题(7a)有封闭解:

$$\mathbf{Y}_{k+1} = (\beta\mu\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{B} + \beta\mu\mathbf{X}_k - \mu\mathbf{A}_k). \quad (8)$$

其中, $(\mathbf{A})^{-1}$ 表示矩阵 \mathbf{A} 的逆矩阵。当 $m \leq n$ 时, 利用 Sherman-Morrison-Woodbury 矩阵逆引理^[6] 可得 $(\beta\mu\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1} = \mathbf{I} - \mathbf{A}^T(\beta\mu\mathbf{I} + \mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}$, 于是(8)式就可以改写为

$$\mathbf{Y}_{k+1} = (\mathbf{I} - \mathbf{A}^T(\beta\mu\mathbf{I} + \mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})(\mathbf{A}^T\mathbf{B} + \beta\mu\mathbf{X}_k - \mu\mathbf{A}_k), \quad (9)$$

这样就可以将一个 $n \times n$ 的矩阵求逆转化成一个 $m \times m$ 的矩阵求逆。

子问题(7b)等价于

$$\mathbf{X}_{k+1} = \arg \min_{\mathbf{X}} \left\{ \log \det(\mathbf{I} + \mathbf{X}^T\mathbf{X}) + \frac{\beta}{2} \left\| \mathbf{X} - \mathbf{Y}_{k+1} - \frac{1}{\beta}\mathbf{A}_k \right\|_F^2 \right\}. \quad (10)$$

此问题同样有封闭解, 下面给出求解此问题的几个定理。

定理 1^[7,8] 对于绝对对称函数 g 和函数 $G(\mathbf{X}) = g(\sigma(\mathbf{X})) = g\sigma(\mathbf{X})$, 其中 $\mathbf{X} \in \mathbb{R}^{m \times n}$ 有奇异值分解(singular values decomposition, SVD) $\mathbf{X} = \mathbf{U} \text{diag}(\sigma(\mathbf{X})) \mathbf{V}^T$ 。假设函数 g 是可微的, 那么 $G(\mathbf{X})$ 的梯度是 $\nabla_{\mathbf{X}} G(\mathbf{X}) = \mathbf{U} \text{diag}(\theta) \mathbf{V}^T$, 其中 $\theta = \nabla_y g(y)|_{y=\sigma(\mathbf{X})}$ 。

定理 2^[7] 假设 $G(\mathbf{X})$ 是酉不变的, 并且矩阵 \mathbf{D}_k 有奇异值分解 $\mathbf{D}_k = \mathbf{U} \text{diag}(\sigma(\mathbf{D}_k)) \mathbf{V}^T$, 那么优化问题 $\min_{\mathbf{X}} \left\{ G(\mathbf{X}) + \frac{\beta}{2} \left\| \mathbf{X} - \mathbf{D}_k \right\|_F^2 \right\}$ 的解是 $\mathbf{X}^* = \mathbf{U} \text{diag}(\sigma^*) \mathbf{V}^T$, 其中 $\sigma^* = \text{prox}_{g,\beta}(\sigma(\mathbf{D}_k))$, 并且 $\text{prox}_{g,\beta}(v) = \arg \min_x \left\{ g(x) + \frac{\beta}{2} \|x - v\|_2^2 \right\}$ 。

因为 $G(\mathbf{X}) = g\sigma(\mathbf{X})$, 且 $g(\sigma) = \sum_i \log(1 + \sigma_i^2)$ 是绝对对称和可微的, 因此易证函数 $G(\mathbf{X}) = \log \det(\mathbf{I} + \mathbf{X}^T\mathbf{X})$ 是酉不变, 可微的。注意到函数 $g(\sigma)$ 关于每个 σ_i 是可分离的, 由定理 2, 问题(10)可转化成下面的标量最小化问题:

$$\arg \min_{\sigma_i} \{ g_{\mathbf{X}, \mathbf{D}_k}(\sigma_i) \} \text{ s. t. } \sigma_i \geq 0, \quad (11)$$

其中: $\mathbf{D}_k = \mathbf{Y}_{k+1} + \frac{1}{\beta}\mathbf{A}_k$, $g_{\mathbf{X}, \mathbf{D}_k}(\sigma_i) = \frac{\beta}{2}(\sigma_i - \sigma_i^{D_k})^2 + \log(1 + \sigma_i^2)$, 且 $\sigma_i^{D_k}$ 是矩阵 \mathbf{D}_k 的奇异值。由一阶最优性条件,

$$\beta\sigma_i^3 - \beta\sigma_i^{D_k}\sigma_i^2 + (\beta + 2)\sigma_i - \beta\sigma_i^{D_k} = 0 \text{ s. t. } \sigma_i \geq 0. \quad (12)$$

性质 1^[7] 设函数 $g(x) = \log(1 + x^2)$, 如果 $\sigma_i^{D_k} = 0$, 则 $\text{prox}_{g,\beta}(\sigma_i^{D_k}) = 0$; 如果 $\sigma_i^{D_k} = 0$ 且 $\beta > \frac{1}{4}$, 则 $\text{prox}_{g,\beta}(\sigma_i^{D_k})$ 落在区间 $(0, \sigma_i^{D_k})$ 内, 且是问题(12)的唯一正根。

由上述结论可得本文的算法框架如下。

算法 1: 求解问题(5)的增广拉格朗日交替方向法(ALADMLD)

输入: $\mu > 0, \beta > \frac{1}{4}$, 矩阵 \mathbf{A}, \mathbf{B} , 最大迭代次数 k_{\max} ;

初始化: 任给矩阵 $\mathbf{X}_0 \in \mathbb{R}^{m \times n}$, $\mathbf{A}_0 \in \mathbb{R}^{m \times n}$, 置迭代次数 $k = 0$;

步骤 1: 根据式(8)或者(9)计算 \mathbf{Y}_{k+1} ;

步骤 2: 根据(10)和性质 1 计算 \mathbf{X}_{k+1} ;

步骤 3: 计算 $\mathbf{A}_{k+1} = \mathbf{A}_k - \gamma\beta(\mathbf{X}_{k+1} - \mathbf{Y}_{k+1})$, $\gamma \in \left(0, \frac{\sqrt{5} + 1}{2}\right)$;

直到 $k > k_{\max}$ 或者 $\{\mathbf{X}_k, \mathbf{Y}_k, \mathbf{A}_k\}$ 收敛;

输出: $\mathbf{X}^* = \mathbf{X}_k$ 。

备注 1: 算法 ALADMLD 步骤 3 中 γ 是为了加快算法的收敛速度。本文的实例中取 $\gamma = 1.5$ 。

3 算法的收敛性分析

本节中, 我们将证明本文所提出的算法收敛到原问题的稳定点。

问题(5)的 KKT 条件为:

$$\frac{1}{\mu} \mathbf{A}^T (\mathbf{A}\mathbf{Y} - \mathbf{B}) = -\mathbf{\Lambda}, \mathbf{X} - \mathbf{Y} = 0, \nabla_{\mathbf{X}} G(\mathbf{X}) = \mathbf{\Lambda}. \quad (13)$$

引理 1 对于 $\beta > \frac{1}{4}$ 和任意初始点 $\mathbf{X}_0 \in \mathbf{R}^{m \times n}, \mathbf{A}_0 \in \mathbf{R}^{m \times n}$, 假设 $\{(\mathbf{X}_k, \mathbf{Y}_k, \mathbf{\Lambda}_k)\}$ 是由 ALADMLD 算法产生的迭代序列, 则序列 $\{\mathbf{\Lambda}_k\}$ 是有界的。

证明: 子问题(7b)的最优解 \mathbf{X}_{k+1} 必须满足一阶最优性条件

$$\nabla_{\mathbf{X}} L_{\beta}(\mathbf{X}, \mathbf{Y}_{k+1}; \mathbf{\Lambda}_k) |_{\mathbf{X}_{k+1}} = \nabla_{\mathbf{X}} G(\mathbf{X}) |_{\mathbf{X}_{k+1}} + \beta(\mathbf{X}_{k+1} - \mathbf{Y}_{k+1}) - \mathbf{\Lambda}_k = 0. \quad (14)$$

将式(7c)代入(14)式, 得到:

$$\nabla_{\mathbf{X}} G(\mathbf{X}) |_{\mathbf{X}_{k+1}} = \mathbf{\Lambda}_{k+1}. \quad (15)$$

由定理 1 可得

$$\nabla_{\mathbf{X}} G(\mathbf{X}) |_{\mathbf{X}_{k+1}} = \mathbf{U} \text{diag} \left(\frac{2\sigma_1}{1 + \sigma_1^2}, \dots, \frac{2\sigma_{n_1}}{1 + \sigma_{n_1}^2} \right) \mathbf{V}^T, \quad (16)$$

其中, $n_1 = \min\{m, n\}$ 且 $0 \leq \frac{2\sigma_i}{1 + \sigma_i^2} \leq 1$, 于是 $\nabla_{\mathbf{X}} G(\mathbf{X}) |_{\mathbf{X}_{k+1}}$ 有界, 从而序列 $\{\mathbf{\Lambda}_k\}$ 有界。得证。

引理 2 对于 $\beta > 1$ 和任意初始点 $\mathbf{X}_0 \in \mathbf{R}^{m \times n}, \mathbf{A}_0 \in \mathbf{R}^{m \times n}$, 假设 $\{(\mathbf{X}_k, \mathbf{Y}_k, \mathbf{\Lambda}_k)\}$ 是由 ALADMLD 算法产生的迭代序列, 则: ① $L_{\beta}(\mathbf{X}_{k+1}, \mathbf{Y}_{k+1}; \mathbf{\Lambda}_k)$ 有界; ② 序列 $\{\mathbf{X}_k\}$ 和 $\{\mathbf{Y}_k\}$ 都是有界的。

证明: 1) 由增广拉格朗日函数的定义(6)及算法步骤可得

$$L_{\beta}(\mathbf{X}_k, \mathbf{Y}_k; \mathbf{\Lambda}_k) = L_{\beta}(\mathbf{X}_k, \mathbf{Y}_k; \mathbf{\Lambda}_{k-1}) + [\mathbf{\Lambda}_{k-1} - \mathbf{\Lambda}_k, \mathbf{X}_k - \mathbf{Y}_k] = L_{\beta}(\mathbf{X}_k, \mathbf{Y}_k; \mathbf{\Lambda}_{k-1}) + \frac{1}{\beta} \|\mathbf{\Lambda}_k - \mathbf{\Lambda}_{k-1}\|_F^2$$

上式的最后一个等式由(7c)得出。因此

$$\begin{aligned} L_{\beta}(\mathbf{X}_{k+1}, \mathbf{Y}_{k+1}; \mathbf{\Lambda}_k) &\leq L_{\beta}(\mathbf{X}_k, \mathbf{Y}_{k+1}; \mathbf{\Lambda}_k) \leq L_{\beta}(\mathbf{X}_k, \mathbf{Y}_k; \mathbf{\Lambda}_k) \leq \\ &L_{\beta}(\mathbf{X}_k, \mathbf{Y}_k; \mathbf{\Lambda}_{k-1}) + \frac{1}{\beta} \|\mathbf{\Lambda}_k - \mathbf{\Lambda}_{k-1}\|_F^2 \leq \dots \leq \\ &L_{\beta}(\mathbf{X}_1, \mathbf{Y}_1; \mathbf{\Lambda}_0) + \sum_{i=1}^k \frac{1}{\beta} \|\mathbf{\Lambda}_i - \mathbf{\Lambda}_{i-1}\|_F^2. \end{aligned}$$

由引理 1, 序列 $\{\mathbf{\Lambda}_k\}$ 有界, 所以上述最后一个不等式的第二项是有限的, 因此 $L_{\beta}(\mathbf{X}_{k+1}, \mathbf{Y}_{k+1}; \mathbf{\Lambda}_k)$ 是有界的。

$$2) \text{ 由 } L_{\beta}(\mathbf{X}_{k+1}, \mathbf{Y}_{k+1}; \mathbf{\Lambda}_k) + \frac{1}{2\beta} \|\mathbf{\Lambda}_k\|_F^2 = G(\mathbf{X}_{k+1}) + \frac{1}{2\mu} \|\mathbf{A}\mathbf{Y}_{k+1} - \mathbf{B}\|_F^2 + \frac{\beta}{2} \|\mathbf{X}_{k+1} - \mathbf{Y}_{k+1} - \frac{1}{\beta} \mathbf{\Lambda}_k\|_F^2 \quad (17)$$

及 $L_{\beta}(\mathbf{X}_{k+1}, \mathbf{Y}_{k+1}; \mathbf{\Lambda}_k)$ 和 $\{\mathbf{\Lambda}_k\}$ 的有界性, 且式(17)右侧各项都是非负的, 可得(17)式的每一项都是有界的。

而 $G(\mathbf{X}_{k+1}) = \sum_i \log(1 + \sigma_i^2(\mathbf{X}_{k+1}))$ 有界, 表明 \mathbf{X}_{k+1} 的所有奇异值都是有界的, 因此序列 $\{\mathbf{X}_k\}$ 是有界的。

再由式(7c) $\frac{1}{\beta} (\mathbf{\Lambda}_k - \mathbf{\Lambda}_{k+1}) = \mathbf{X}_{k+1} - \mathbf{Y}_{k+1}$, 即可得序列 $\{\mathbf{Y}_k\}$ 的有界性。得证。

定理 3 对于 $\beta > 1$ 和任意初始点 $\mathbf{X}_0 \in \mathbf{R}^{m \times n}, \mathbf{A}_0 \in \mathbf{R}^{m \times n}$, 假设 $\{\mathbf{X}_k, \mathbf{Y}_k, \mathbf{\Lambda}_k\}$ 是由 ALADMLD 算法产生的迭代序列, 则序列 $\{\mathbf{X}_k, \mathbf{Y}_k, \mathbf{\Lambda}_k\}$ 至少有一个聚点 $\{\mathbf{X}^*, \mathbf{Y}^*, \mathbf{\Lambda}^*\}$, 且 $\{\mathbf{X}^*, \mathbf{Y}^*\}$ 是问题(4)的一个稳定点。

证明: 由引理 1 和引理 2 可知, 序列 $\{\mathbf{X}_k, \mathbf{Y}_k, \mathbf{\Lambda}_k\}$ 有界。根据 Bolzano-Weierstrass 定理, 序列 $\{\mathbf{X}_k, \mathbf{Y}_k, \mathbf{\Lambda}_k\}$ 至少存在一个聚点, 记为 $\{\mathbf{X}^*, \mathbf{Y}^*, \mathbf{\Lambda}^*\}$ 。不失一般性, 假设 $\{\mathbf{X}_k, \mathbf{Y}_k, \mathbf{\Lambda}_k\}$ 收敛到 $\{\mathbf{X}^*, \mathbf{Y}^*, \mathbf{\Lambda}^*\}$ 。

由式(7c)可得 $\frac{1}{\beta} (\mathbf{\Lambda}_{k-1} - \mathbf{\Lambda}_k) = \mathbf{X}_k - \mathbf{Y}_k$, 再由 $\beta > 1$ 和序列 $\{\mathbf{\Lambda}_k\}$ 有界可得, 当 $k \rightarrow \infty$ 时, 有

$$\mathbf{X}^* - \mathbf{Y}^* = 0. \quad (18)$$

当 $k \rightarrow \infty$, 由(14), (15)式得

$$\nabla_{\mathbf{X}} G(\mathbf{X}) |_{\mathbf{X}^*} = \mathbf{\Lambda}^*. \quad (19)$$

由(7a)式的一阶最优性条件, 可得

$$\frac{1}{\mu} \mathbf{A}^T (\mathbf{A} \mathbf{Y}_k - \mathbf{B}) + \beta (\mathbf{Y}_k - \mathbf{X}_{k-1}) + \mathbf{\Lambda}_{k-1} = 0. \quad (20)$$

将(7c)代入(20)式得到

$$\frac{1}{\mu} \mathbf{A}^T (\mathbf{A} \mathbf{Y}_k - \mathbf{B}) = \beta (\mathbf{X}_{k-1} - \mathbf{X}_k) - \mathbf{\Lambda}_k. \quad (21)$$

由序列 $\{\mathbf{X}_k\}$ 有界且 $\beta > 1$ 可得, 当 $k \rightarrow \infty$ 时, 有

$$\frac{1}{\mu} \mathbf{A}^T (\mathbf{A} \mathbf{Y}^* - \mathbf{B}) = -\mathbf{\Lambda}^*. \quad (22)$$

由(18),(19)和(22)式可知, $\{\mathbf{X}^*, \mathbf{Y}^*, \mathbf{\Lambda}^*\}$ 是问题(5)的一个稳定点。得证。

4 实验结果与分析

本节对已有的求解核范数矩阵秩最小化问题的较先进的算法—APGL^[1], LADM^[3], LSA^[10], SCC^[11], LRR^[12], LRSC^[13], SSC^[14]与本文提出的 ALADMLD 算法进行比较。需要指出的是 APGL 代码可以从软件包 SLEP 中获得。本文所有的实验都是在 Windows 8 系统 MATLABR2013a 中运行的。

实验分为三类:

- 1) 4.1 节采用实际数据(Scene 和股票数据), 比较 APGL 与 ALADMLD 求解多元线性分析系数问题;
- 2) 4.2 节采用随机数据, 比较 LADM 与 ALADMLD 算法的高效性;
- 3) 4.3 节采用 Extended Yale B 数据应用到人脸识别, 比较 ALADMLD 算法与 LSA, SCC, LRR, LRSC, SSC 的有效性。

由于算法中参数比较多, 首先给出实验的基本细节。

1) 近似参数 τ 。在 APGL 中, 取 $\tau_0 = 100$, 在 LADM 中, $\tau = \frac{1}{\rho(\mathbf{A}^T \mathbf{A})}$, 其中 $\rho(\mathbf{A}^T \mathbf{A})$ 表示矩阵 $\mathbf{A}^T \mathbf{A}$ 的谱半径。

2) 罚参数 β 。在 LADM 中, $\beta = 2.5 / \min(m, n)$, 在 ALADMLD 算法中, 取 $\beta_0 = 1$ 。

3) 参数 μ 。为了验证算法 ALADMLD 的有效性, 实验采用多个 μ 值进行测试, 如表 1 所示。

4) 终止准则。ALADMLD 和 LADM 算法的终止准则:

$$XChag = \frac{\|\mathbf{X}_k - \mathbf{X}_{k-1}\|_F}{\max(\|\mathbf{X}_{k-1}\|_F, 1)} \leq tol,$$

其中, $tol = 10^{-6}$ 是终止参数。实验的最大迭代次数 $k_{max} = 100$ 。

5) 评价标准。均方根误差(root mean square error, RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (\tilde{y}_i - y_i)^2}{m}}$$

表 1 给出了所有实验数据的统计信息。

表 1 实验中的数据集统计信息

Tab. 1 Statistics for the data sets used in our tests

编号	p	m	n	μ	数据名称
1	50	294	6	10^{-4}	Scene
2	800	3 724	800	10^{-4}	Stock
3	2 500	2 000	100	$10^{-2}, 10^{-1}$	Rand
4	2 500	2 000	100	$10^{-2}, 10^{-1}$	Rand
5	38	—	—	19	EYaleB

4.1 APGL 与 ALADMLD 的实验结果

本小节采用两种实际数据: Scene 和 Stock, 这些数据都可以从网上下载。为了使算法具有可比性, 本实验采用的都是 APGL 算法使用的默认参数。图 2 给出了算法 APGL 和 ALADMLD 利用两种实际数据计算

得出的均方根误差随迭代次数的变化。图 2(a)中算法 APGL 和 ALADMLD 的迭代次数分别是 100 和 5, 图 2(b)中算法 APGL 和 ALADMLD 的迭代次数分别是 6 和 5。虽然图 2(b)中的迭代次数很相近,但是就均方根误差来说,算法 ALADMLD 的精度比 APGL 高出很多。

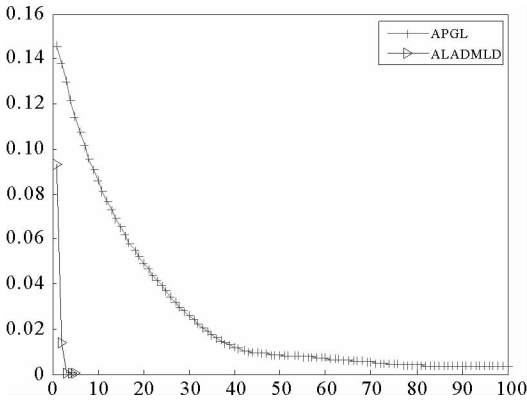


图 2(a) 测试 Scene 数据的均方根误差
Fig. 2(a) Test of RMSE on Scene data

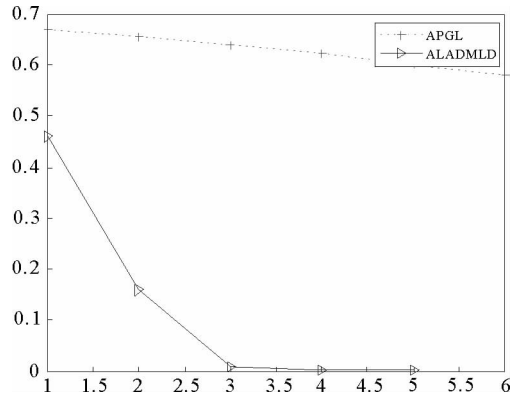


图 2(b) 测试 Stock 数据的均方根误差
Fig. 2(b) Test of RMSE on Stock data

4.2 LADM 与 ALADMLD 的实验结果

因为算法 LADM 与 ALADMLD 都属于增广拉格朗日方法,故两种算法的比较更具有说服力。为了更好地验证算法的有效性,此实验采用文献[3]的方式生成数据矩阵 \mathbf{A} ,具体产生过程如下:

- 1) $\text{sqrt_lam} = \text{sqrt}(\text{lam_max});$
- 2) $[\mathbf{P}, \sim] = \text{qr}(\text{randn}(p));$
- 3) $d = [1; 1 + \text{rand}(\text{min}(p, m) - 2, 1) * (\text{sqrt_lam} - 1); \text{sqrt_lam}];$
- 4) $[\mathbf{Q}, \sim] = \text{qr}(\text{randn}(m));$
- 5) $\mathbf{A} = \mathbf{P} * \text{spdiags}(d, 0, p, m) * \mathbf{Q}'.$

其中, lam_max 为矩阵 $\mathbf{A}\mathbf{A}^T$ 的最大特征值。利用 $\mathbf{X}_{op} = \text{randn}(m, r) * \text{randn}(r, n)$ 和 $\mathbf{B} = \mathbf{A}\mathbf{X}_{op} + \text{std} * \text{randn}(p, n)$ 产生低秩矩阵 \mathbf{X}_{op} 和 \mathbf{B} , 其中 $\text{std} = 0.001$ 。实验结果如表 2 所示,其中相对误差 $XRel =$

$$\frac{\|\mathbf{X}_k - \mathbf{X}_{op}\|_F}{\|\mathbf{X}_{op}\|_F}, \text{Iter}, r \text{ 和 } \text{Time} \text{ 分别表示迭代次数、低秩矩阵 } \mathbf{X}_{op} \text{ 的秩和计算时间(单位是秒)。表 2 中的}$$

2.7-2 表示 2.7×10^{-2} 。由表 2 可以得出,随着矩阵维数的增大,两种算法都会消耗更多的 CPU 时间,但是 LADM 比 ALADMLD 算法消耗的 CPU 时间更多,且对于高维数据,ALADMLD 的 CPU 时间一般在 30 秒内终止;就迭代次数而言,ALADMLD 算法的迭代次数不超过 10 次,且波动不大;就均方根误差而言,ALADMLD 算法比 LADM 算法更小。基于以上分析,对于各项评价指标,ALADMLD 算法都优于 LADM。因此,随着大数据的应用,ALADMLD 算法将比 LADM 算法具有更广阔的应用前景。

4.3 LSA, SCC, LRR, SSC, LRSC 与 ALADMLD 的实验结果

由于低秩表示方法在子空间聚类中也有应用^[12],故本小节将 ALADMLD 算法应用到人脸识别,利用 Extended Yale B 数据验证算法的有效性。该数据库包括 38 个样本,每个样本有 64 张在不同光线下拍摄的前额照片。借鉴文献[7, 13, 14]的处理方式,降低采样照片像素至 48×42 ,然后将每张图片向量化为一个数据点。将 38 个样本划分为 4 组:1~10, 11~20, 21~30, 31~38。针对每组依次选取 $n \in \{2, 3, 5, 8, 10\}$ 个样本。表 3 给出了各种算法的聚类误差,此处聚类误差定义为聚类错误的数据点总和与总的数据点的比值。由表 3 可以看出,当样本数目是 2 或者 3 时,ALADMLD 算法的性能比 SSC 稍微逊色,而在其余所有的情景下,ALADMLD 算法比其他几个算法都准确。另外,需要指出的是,ALADMLD 算法更适合多目标聚类问题,这恰恰迎合了大数据时代对多目标聚类问题的基本要求。

表 2 算法 LADM 和 ALADMLD 的结果比较

Tab. 2 Results comparison of LADM and ALADMLD

参数					LADM				ALADMLD			
(λ, μ)	p	m	n	r	XRel	Iter	Time	RMSE	XRel	Iter	Time	RMSE
(1,10-2)	2 500	2 000	100	20	2.7-2	100	34.5	0.367 3	7.1-4	4	5.87	0.005 0
	3 000	3 000	200	50	1.5-1	100	116.3	3.065 7	6.6-4	4	16.21	0.001 0
(1,10-1)	2 500	2 000	100	20	5.1-2	100	34	0.698 3	6.4-4	8	9.87	0.005 4
	3 000	3 000	200	50	1.8-1	100	108.2	3.729 3	6-4	8	29.67	0.003 4
(10,10-2)	2 500	2 000	100	20	3.8-4	69	23.3	0.011 6	3.8-4	4	5	0.005 1
	3 000	3 000	200	50	4.9-4	63	70.4	0.015 8	3.5-4	3	15	0.002 2
(10,10-1)	2 500	2 000	100	20	9.5-4	69	23.3	0.023 5	3.6-4	6	7.9	0.005 1
	3 000	3 000	200	50	1.7-3	63	74.1	0.053 9	3.5-4	6	25.3	0.001 6
(102,10-2)	2 500	2 000	100	20	不收敛				2.1-4	3	4.7	0.005 2
	3 000	3 000	200	50	5.5-1	100	116.3	16.05	2.1-4	3	14.9	0.002 2
(102,10-1)	2 500	2 000	100	20	不收敛				2.1-4	6	8.0	0.005 0
	3 000	3 000	200	50	5.2-1	100	111	15.17	2.1-4	5	20.1	0.001 8

4.4 实验总结

通过实验结果的比较发现,本文提出的算法 ALADMLD 算法在许多方面较现有的求解核范数的有效算法—APGL 和 LADM 更高效,具体可归结为以下几点:

- 1) 三种算法都属于一阶方法,即在每次迭代过程中,只需要计算函数值和梯度信息,因此三种算法都可用于求解大规模数据问题,但是当终止准则相同时,ALADMLD 的计算速度更快;
- 2) APGL 和 LADM 算法的有效性严格依赖于近似映射,近似参数 τ 的选取不容易把握;而 ALADMLD 算法则不需要对子问题进行近似映射,不涉及近似参数 τ 的选取,因此更加高效。
- 3) 当矩阵的某些奇异值比较大时,采用对数行列式函数近似矩阵秩函数的近似效果优于核范数的近似效果。

表 3 应用 Extended Yale B 数据的聚类误差

Tab. 3 Clustering error rates (%) on Extended Yale B

个体数	方法	LSA	SCC	LRR	LRSC	SSC	ALADMLD
2	平均值	32.80	16.62	9.52	5.32	1.86	2.720
	中位数	47.66	7.82	5.47	4.69	0.00	0.078
3	平均值	52.29	38.16	19.52	8.47	3.10	3.640
	中位数	50.00	39.06	14.58	7.81	1.04	1.506
5	平均值	58.02	58.90	34.16	12.24	4.31	4.600
	中位数	56.87	59.38	35.00	11.25	2.50	2.500
8	平均值	59.19	66.11	41.19	23.72	5.85	4.930
	中位数	58.59	64.65	43.75	28.03	4.49	3.610
10	平均值	60.42	73.02	38.85	30.36	10.94	6.300
	中位数	57.50	75.78	41.09	28.75	5.63	4.840

5 结论

针对矩阵秩最小化问题,提出了对数行列式函数作为矩阵秩函数的非凸近似。虽然此矩阵秩最小化问

题是非凸的,通过引入辅助变量,应用增广拉格朗日交替方向法求解此矩阵秩最小化问题。当罚参数 $\beta > 1$ 时,证明了提出的算法—ALADMLD 产生的序列收敛到非凸矩阵秩最小化问题的稳定点。利用随机数据和实际数据,将 ALADMLD 算法与基于核范数的先进算法进行数值实验比较,验证提出的算法高效性,同时表明了随着大数据问题的引入与推广,ALADMLD 算法将比现有的 LADM 算法具有更广阔的应用前景。

参考文献:

- [1]TOH K, YUN S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems[J]. Pacific Journal of Optimization, 2010, 6(615-640): 15.
- [2]LIN Z, CHEN M, MA Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices[J]. arXiv preprint arXiv:1009.5055, 2010.
- [3]YANG J, YUAN X. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization[J]. Mathematics of Computation, 2013, 82(281): 301-329.
- [4]CHEN C, HE B, YUAN X. Matrix completion via an alternating direction method[J]. IMA Journal of Numerical Analysis, 2012, 32(1): 227-245.
- [5]DUDIK M, HARCHAOU I, MALICK J. Lifted coordinate descent for learning with trace-norm regularization[C]//AIST-ATS-Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics-2012. 2012, 22: 327-336.
- [6]SHERMAN J, MORRISON W. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix [J]. The Annals of Mathematical Statistics, 1950: 124-127.
- [7]PENG C, KANG Z, LI H, et al. Subspace clustering using log-determinant rank approximation[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 925-934.
- [8]LEWIS A, SENDOV H. Nonsmooth analysis of singular values; Part I: Theory[J]. Set-Valued Analysis, 2005, 13(3): 213-241.
- [9]SHEN Y, WEN Z, ZHANG Y. Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization[J]. Optimization Methods and Software, 2014, 29(2): 239-263.
- [10]YAN J, POLLEFEYS M. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate[M]//Computer Vision/CECCV 2006. Springer Berlin Heidelberg, 2006: 94-106.
- [11]CHEN G, LERMAN G. Spectral curvature clustering (SCC)[J]. International Journal of Computer Vision, 2009, 81(3): 317-330.
- [12]LIU G, LIN Z, YU Y. Robust subspace segmentation by low-rank representation[C]//Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 663-670.
- [13]VIDAL R, FAVARO P. Low rank subspace clustering (LRSC)[J]. Pattern Recognition Letters, 2014, 43: 47-61.
- [14]ELHAMIFAR E, VIDAL R. Sparse subspace clustering: Algorithm, theory, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2765-2781.
- [15]FAZEL M. Matrix rank minimization with applications[D]. Palo Alto: Stanford University, 2002.

(责任编辑:傅游)