

求解一类矩阵迹极小化问题的非线性共轭梯度法

李春梅^{1,2}, 王翠方², 段雪峰²

(1. 贵州师范大学数学科学学院, 贵州 贵阳 550001)

(2. 桂林电子科技大学数学与计算科学学院, 广西 桂林 541004)

摘要: 本文研究了图分割问题中的矩阵迹极小化问题. 利用半正定矩阵的 Gramian 表示, 将该问题转化为无约束优化问题, 设计了 Armijo 线搜索下的非线性共轭梯度方法进行求解. 数值例子表明新方法是可行的.

关键词: 矩阵迹极小化; Gramian 表示; 非线性共轭梯度法

MR(2010) 主题分类号: 15A15; 49M99

中图分类号: 0241.6

文献标识码: A

文章编号: 0255-7797(2020)03-0323-09

1 引言

本文用 $\mathbf{R}^{n \times n}$ 和 $\mathbf{SR}_+^{n \times n}$ 分别表示 $n \times n$ 阶实矩阵和 $n \times n$ 阶实对称半正定矩阵构成的集合. 用 $\text{diag}(\mathbf{Y})$ 表示矩阵 $\mathbf{Y}$ 的对角线元素组成的向量. 用 $\text{tr}(\mathbf{Y})$, $\mathbf{Y}^T$ 和 $\text{rank}(\mathbf{Y})$ 分别表示矩阵的迹, 转置和秩. $\mathbf{e}$ 表示全 1 向量. 本文主要研究图分割中的矩阵迹极小化问题

$$\min \text{tr}(\mathbf{QY}), \quad \text{s.t. } \text{diag}(\mathbf{Y}) = \mathbf{e}, \quad \mathbf{Y} \in \mathbf{SR}_+^{n \times n}. \quad (1.1)$$

该问题来自于加权无向图的最小割问题, 具体如下: 给定一个加权无向图 $G = (V, E)$, 最小割问题在于寻找一种节点的划分, 使得切边的权重之和最小^[1], 即

$$\min \mathbf{y}^T \mathbf{Q} \mathbf{y}, \quad \text{s.t. } y_i^2 = 1, i = 1, \dots, n, \quad (1.2)$$

其中 $\mathbf{Q} = \frac{1}{4}(\text{diag}(\mathbf{A}\mathbf{e}) - \mathbf{A})$ 是该图的负 Laplace 矩阵, $\mathbf{A}$ 为加权无向图的邻接矩阵. 又对于任意的对称矩阵 $\mathbf{M}$ 都可以表示为图的 Laplace 矩阵加上对角阵 $\mathbf{D}$ 的形式, 即 $\mathbf{M} = \mathbf{Q} + \mathbf{D}$, 且对于任意的 $\mathbf{y} \in \{-1, 1\}^n$ 都有 $\mathbf{y}^T \mathbf{D} \mathbf{y} = \text{tr}(\mathbf{D})$, 所以问题 (1.2) 等价于如下问题.

$$\min \mathbf{y}^T \mathbf{M} \mathbf{y}, \quad \text{s.t. } \mathbf{y} \in \{-1, 1\}^n. \quad (1.3)$$

令 $\mathbf{Y} = \mathbf{y}\mathbf{y}^T$, 则问题 (1.3) 可等价写成 (1.1). 另外, 矩阵迹极小化问题还来源于图像还原^[2]、数据降维^[3]、机器学习^[4] 等领域.

近年来, 许多人研究了矩阵迹极小化问题, 许多数值方法被提出. 比如, 1992 年, Henk^[5] 等提出了一类矩阵迹极小化的算法, 它包含了 Procrustes rotation 加权算法, 有效的提高

*收稿日期: 2018-06-30 接收日期: 2019-04-26

基金项目: 国家自然科学基金项目 (11561015; 11761024); 广西科技项目 (AD18281024); 广西自然科学基金项目 (2016GXNSFFA380009; 2017GXNSFBA198082; 2016XNSFAA380074).

作者简介: 李春梅 (1984-), 女, 湖南邵阳, 副研究员, 主要研究方向: 数值代数.

了图像分割中的矩阵迹极小化问题的计算效率. 1996 年, Bental^[6] 等人实现了由双边约束下的不确定二次矩阵迹极小化问题到简单线约束的凸矩阵迹极小化问题的转化. 2000 年, Benson^[7] 等提出了一个双尺度内点法, 解决了布尔约束条件下组合矩阵迹极小化问题和二次矩阵迹极小化问题的半正定松弛问题. 2009 年, Burer^[8] 等提出了二级松弛, 解决了计算量与矩阵规模不平衡的问题, 从另一个层面上提高了解决迹函数极小化的计算效率. 随后, 2011 年 Grippo^[1] 等人提出一个连续可微的精确价值函数, 并利用该价值函数设计了具有全局收敛性的低秩半正定松弛的精确罚算法, 从而解决了图像分割中的二次等式约束的矩阵迹极小化问题. 2016 年, Liu^[9] 等人设计了一个新的 Riemannian 信赖域方法来解决 Stiefel 流形上的矩阵迹极大化问题, 即相对于典型的 Riemannian 信赖域方法提出了交替方向法以及高斯-赛德尔方法的初值策略, 这一方法在收敛速度和节省储存量方面有明显优势. 然而, 对于求解问题 (1.1) 的数值方法较少.

本文利用 Gramian 表示和三角变换将矩阵迹极小化问题 (1.1) 转化为无约束优化问题. 然后设计了 Armijo 线搜索下的非线性共轭梯度法求解无约束优化问题. 最后用数值例子验证了新方法的可行性.

2 主要结果

本节先利用 Gramian 表示刻画问题 (1.1) 的可行集, 将该问题转化为无约束优化问题, 再设计非线性共轭梯度法求解. 先给出两个引理.

引理 2.1 (Gramian 表示)^[10] 设 Y 为 $n \times n$ 阶对称半正定矩阵, 则 $\text{rank}(Y) \leq r$ 当且仅当存在一个矩阵 $V \in \mathbf{R}^{n \times k}$ 满足 $Y = VV^T$.

由引理 2.1 知可利用 $Y = VV^T, V \in \mathbf{R}^{n \times r}$ 刻画问题 (1.1) 的可行集, 即

$$\Omega = \{Y | Y \in \mathbf{SR}_+^{n \times n}, \text{rank}(Y) \leq r\}.$$

因此问题 (1.1) 可等价转化为下述优化问题:

$$\min \text{tr}(QVV^T), \quad \text{s.t. } \text{diag}(VV^T) = e, \quad V \in \mathbf{R}^{n \times r}. \quad (2.1)$$

为了刻画问题 (2.1) 的可行集, 引入如下引理.

引理 2.2^[11] 令矩阵 $V = [V_1, V_2, \dots, V_r] = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1r} \\ v_{21} & v_{22} & \cdots & v_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nr} \end{bmatrix} \in \mathbf{R}^{n \times r}$, 设

$$V_1 = \begin{bmatrix} \cos \alpha_{11} \\ \cos \alpha_{21} \\ \vdots \\ \cos \alpha_{n1} \end{bmatrix}, \quad V_2 = \begin{bmatrix} \cos \alpha_{11} \sin \alpha_{11} \\ \cos \alpha_{21} \sin \alpha_{21} \\ \vdots \\ \cos \alpha_{n1} \sin \alpha_{n1} \end{bmatrix}, \quad \dots, \quad V_{r-1} = \begin{bmatrix} \cos \alpha_{1,r-1} \prod_{l=1}^{r-2} \sin \alpha_{1l} \\ \cos \alpha_{2,r-1} \prod_{l=1}^{r-2} \sin \alpha_{2l} \\ \vdots \\ \cos \alpha_{n,r-1} \prod_{l=1}^{r-2} \sin \alpha_{nl} \end{bmatrix},$$

$$V_r = \begin{bmatrix} \prod_{l=1}^{r-1} \sin \alpha_{1l} \\ \prod_{l=1}^{r-1} \sin \alpha_{2l} \\ \vdots \\ \prod_{l=1}^{r-1} \sin \alpha_{nl} \end{bmatrix},$$

其中 $\alpha_{ij} \in \mathbf{R}, i = 1, 2, \dots, r-1$, 则 $\mathbf{Y} = \mathbf{V}\mathbf{V}^T \in \mathbf{SR}_+^{n \times n}$, 且 $\text{rank}(\mathbf{Y}) \leq r, \text{diag}(\mathbf{Y}) = \mathbf{e}$.

注 2.1 引用一个简单的 3×2 阶矩阵解释引理 2.2. 令

$$\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2] = \begin{bmatrix} \cos \alpha_{11} & \sin \alpha_{11} \\ \cos \alpha_{21} & \sin \alpha_{21} \\ \cos \alpha_{31} & \sin \alpha_{31} \end{bmatrix}.$$

易知

$$\begin{aligned} \mathbf{Y} = \mathbf{V}\mathbf{V}^T &= \begin{bmatrix} \cos \alpha_{11} & \sin \alpha_{11} \\ \cos \alpha_{21} & \sin \alpha_{21} \\ \cos \alpha_{31} & \sin \alpha_{31} \end{bmatrix} \begin{bmatrix} \cos \alpha_{11} & \sin \alpha_{11} \\ \cos \alpha_{21} & \sin \alpha_{21} \\ \cos \alpha_{31} & \sin \alpha_{31} \end{bmatrix}^T \\ &= \begin{bmatrix} 1 & \cos \alpha_{11} \cos \alpha_{21} + \sin \alpha_{11} \sin \alpha_{21} & \cos \alpha_{11} \cos \alpha_{31} + \sin \alpha_{11} \sin \alpha_{31} \\ \cos \alpha_{11} \cos \alpha_{21} + \sin \alpha_{11} \sin \alpha_{21} & 1 & \cos \alpha_{21} \cos \alpha_{31} + \sin \alpha_{21} \sin \alpha_{31} \\ \cos \alpha_{11} \cos \alpha_{31} + \sin \alpha_{11} \sin \alpha_{31} & \cos \alpha_{21} \cos \alpha_{31} + \sin \alpha_{21} \sin \alpha_{31} & 1 \end{bmatrix}. \end{aligned}$$

显然, 矩阵 $\mathbf{Y}$ 不仅对称半正定, 且满足 $\text{rank}(\mathbf{Y}) \leq 2$, 以及 $\text{diag}(\mathbf{Y}) = \mathbf{e}$.

由引理 2.2 知矩阵 $\mathbf{Y}$ 的每个位置上的元素可以表述为

$$y_{ij} = \begin{cases} \sum_{p=1}^{r-1} \cos \alpha_{ip} \cos \alpha_{jp} \prod_{l=1}^{p-1} \sin \alpha_{il} \sin \alpha_{jl} + \prod_{l=1}^{r-1} \sin \alpha_{il} \sin \alpha_{jl}, & i \neq j, \\ 0, & i = j. \end{cases}$$

由引理 2.2 可将问题 (2.1) 等价转化为如下无约束优化问题

$$\min_{\alpha \in R^{n \times (r-1)}} F(\alpha), \quad (2.2)$$

其中

$$\begin{aligned} F(\alpha) &= \text{tr}(\mathbf{Q}\mathbf{Y}) = \sum_{i=1}^n \sum_{j=1}^n q_{ji} y_{ij} \\ &= \sum_{i=1}^n \sum_{j=1}^n q_{ij} \left(\sum_{p=1}^{r-1} \cos \alpha_{ip} \cos \alpha_{jp} \prod_{l=1}^{p-1} \sin \alpha_{il} \sin \alpha_{jl} + \prod_{l=1}^{r-1} \sin \alpha_{il} \sin \alpha_{jl} \right). \end{aligned}$$

下面将利用非线性共轭梯度法求解问题 (2.2). 先给出目标函数的梯度如下.

定理 2.1 $F(\alpha)$ 的梯度为

$$\nabla F(\alpha) = \left(\frac{\partial F(\alpha)}{\partial \alpha_{11}}, \frac{\partial F(\alpha)}{\partial \alpha_{21}}, \dots, \frac{\partial F(\alpha)}{\partial \alpha_{n1}}, \dots, \frac{\partial F(\alpha)}{\partial \alpha_{1,r-1}}, \frac{\partial F(\alpha)}{\partial \alpha_{2,r-1}}, \dots, \frac{\partial F(\alpha)}{\partial \alpha_{n,r-1}} \right)^T,$$

其中

$$\begin{aligned} & \frac{\partial F(\alpha)}{\partial \alpha_{\mu\nu}} \\ = & \sum_{i=1, i \neq \mu}^n [(q_{\mu i} + q_{j\mu}) (-\sin \alpha_{\mu\nu} \cos \alpha_{i\nu} \prod_{l=1}^{\nu-1} \sin \alpha_{\mu l} \sin \alpha_{il} + \cos \alpha_{\mu\nu} \sin \alpha_{i\nu} \prod_{l=1, l \neq \nu}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{jl} \\ & + \sum_{p=\nu+1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{ip} \cos \alpha_{\mu\nu} \cos \alpha_{i\nu} \prod_{l=1, l \neq \mu}^{p-1} \sin \alpha_{\mu l} \sin \alpha_{il})] \\ & + q_{\mu\mu} (-2 \cos \alpha_{\mu\nu} \sin \alpha_{\mu\nu} \prod_{l=1}^{\nu-1} \sin^2 \alpha_{\mu l} + 2 \sin \alpha_{\mu\nu} \cos \alpha_{\mu l} \prod_{l=1, l \neq \nu}^{r-1} \sin^2 \alpha_{\mu l} \\ & + \sum_{p=\nu+1}^{r-1} 2 \cos \alpha_{\mu p} \sin \alpha_{\mu\nu} \cos \alpha_{\mu\nu} \prod_{l=1, l \neq \mu}^{p-1} \sin^2 \alpha_{\mu l}). \end{aligned}$$

证

$$\begin{aligned} \frac{\partial F(\alpha)}{\partial \alpha_{\mu\nu}} &= \frac{\partial}{\partial \alpha_{\mu\nu}} \left[\sum_{j=1, j \neq \mu}^n q_{j\mu} \left(\sum_{p=1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{jp} \prod_{l=1}^{p-1} \sin \alpha_{\mu l} \sin \alpha_{jl} + \prod_{l=1}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{jl} \right) \right. \\ & \quad + \frac{\partial}{\partial \alpha_{\mu\nu}} \left[\sum_{i=1, i \neq \mu}^n q_{\mu i} \left(\sum_{p=1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{ip} \prod_{l=1}^{p-1} \sin \alpha_{\mu l} \sin \alpha_{il} + \prod_{l=1}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{il} \right) \right. \\ & \quad \left. + \frac{\partial}{\partial \alpha_{\mu\nu}} [q_{\mu\mu} (\sum_{p=1}^{r-1} \cos^2 \alpha_{\mu p} \prod_{l=1}^{p-1} \sin^2 \alpha_{\mu l} + \prod_{l=1}^{r-1} \sin^2 \alpha_{\mu l})] \right] \\ &= \frac{\partial}{\partial \alpha_{\mu\nu}} \left[\sum_{j=1}^{\mu-1} q_{j\mu} \left(\sum_{p=1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{jp} \prod_{l=1}^{p-1} \sin \alpha_{\mu l} \sin \alpha_{jl} + \prod_{l=1}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{jl} \right) \right. \\ & \quad + \sum_{j=\mu+1}^n q_{j\mu} \left(\prod_{p=1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{jp} \prod_{l=1}^{p-1} \sin \alpha_{\mu l} \sin \alpha_{jl} + \prod_{l=1}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{jl} \right) \\ & \quad + \frac{\partial}{\partial \alpha_{\mu\nu}} \left[\sum_{i=1}^{\mu-1} q_{\mu i} \left(\prod_{p=1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{ip} \prod_{l=1}^{p-1} \sin \alpha_{\mu l} \sin \alpha_{il} + \prod_{l=1}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{il} \right) \right. \\ & \quad + \sum_{j=\mu+1}^n q_{\mu i} \left(\prod_{p=1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{ip} \prod_{l=1}^{p-1} \sin \alpha_{\mu l} \sin \alpha_{il} + \prod_{l=1}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{il} \right) \\ & \quad \left. + \frac{\partial}{\partial \alpha_{\mu\nu}} [q_{\mu\mu} (\prod_{p=1}^{r-1} \cos^2 \alpha_{\mu p} \prod_{l=1}^{p-1} \sin^2 \alpha_{\mu l} + \prod_{l=1}^{r-1} \sin^2 \alpha_{\mu l})] \right] \\ &= \left\{ \sum_{j=1}^{\mu-1} [q_{j\mu} (-\sin \alpha_{\mu\nu} \cos \alpha_{j\nu} \prod_{l=1}^{\nu-1} \sin \alpha_{\mu l} \sin \alpha_{jl} + \cos \alpha_{\mu\nu} \sin \alpha_{j\nu} \prod_{l=1, l \neq \nu}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{jl} \right. \end{aligned}$$

$$\begin{aligned}
& + \sum_{p=\nu+1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{j p} \cos \alpha_{\mu \nu} \cos \alpha_{j \nu} \prod_{l=1, l \neq \mu}^{p-1} \sin \alpha_{\mu l} \sin_{j l}]] \\
& + \sum_{j=\mu+1}^n q_{j \mu} [(-\sin \alpha_{\mu \nu} \cos \alpha_{j \nu} \prod_{l=1}^{\nu-1} \sin \alpha_{\mu l} \sin \alpha_{j l} + \cos \alpha_{\mu \nu} \sin \alpha_{j \nu} \prod_{l=1, l \neq \mu}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{j l} \\
& + \sum_{p=\nu+1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{j p} \cos \alpha_{\mu \nu} \cos \alpha_{j \nu} \prod_{l=1, l \neq \mu}^{p-1} \sin \alpha_{\mu l} \sin_{j l}]]\} \\
& + \{ \sum_{i=1}^{\mu-1} [q_{i \mu} (-\sin \alpha_{\mu \nu} \cos \alpha_{i \nu} \prod_{l=1}^{\nu-1} \sin \alpha_{\mu l} \sin \alpha_{i l} + \cos \alpha_{\mu \nu} \sin \alpha_{i \nu} \prod_{l=1, l \neq \nu}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{i l} \\
& + \sum_{p=\nu+1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{i p} \cos \alpha_{\mu \nu} \cos \alpha_{i \nu} \prod_{l=1, l \neq \mu}^{p-1} \sin \alpha_{\mu l} \sin_{i l}]] \\
& + \sum_{i=\mu+1}^n [q_{i \mu} (-\sin \alpha_{\mu \nu} \cos \alpha_{i \nu} \prod_{l=1}^{\nu-1} \sin \alpha_{\mu l} \sin \alpha_{i l} + \cos \alpha_{\mu \nu} \sin \alpha_{i \nu} \prod_{l=1, l \neq \nu}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{i l} \\
& + \sum_{p=\nu+1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{i p} \cos \alpha_{\mu \nu} \cos \alpha_{i \nu} \prod_{l=1, l \neq \mu}^{p-1} \sin \alpha_{\mu l} \sin_{i l}]]\} \\
& + [q_{\mu \mu} (-2 \cos \alpha_{\mu l} \sin \alpha_{\mu \nu} \prod_{l=1}^{\nu-1} \sin^2 \alpha_{\mu l} + 2 \sin \alpha_{\mu \nu} \cos \alpha_{\mu \nu} \prod_{l=1, l \neq \nu}^{r-1} \sin^2 \alpha_{\mu l} \\
& + \sum_{p=\nu+1}^{r-1} 2 \cos^2 \alpha_{\mu p} \sin \alpha_{\mu \nu} \cos \alpha_{\mu \nu} \prod_{l=1, l \neq \mu}^{p-1} \sin^2 \alpha_{\mu l}]] \\
& = \sum_{i=1, i \neq \mu}^n [(q_{\mu i} + q_{j \mu}) (-\sin \alpha_{\mu \nu} \cos \alpha_{i \nu} \prod_{l=1}^{\nu-1} \sin \alpha_{\mu l} \sin \alpha_{i l} + \cos \alpha_{\mu \nu} \sin \alpha_{i \nu} \prod_{l=1, l \neq \nu}^{r-1} \sin \alpha_{\mu l} \sin \alpha_{j l} \\
& + \sum_{p=\nu+1}^{r-1} \cos \alpha_{\mu p} \cos \alpha_{i p} \cos \alpha_{\mu \nu} \cos \alpha_{i \nu} \prod_{l=1, l \neq \mu}^{p-1} \sin \alpha_{\mu l} \sin_{i l}]] \\
& + q_{\mu \mu} (-2 \cos \alpha_{\mu \nu} \sin \alpha_{\mu \nu} \prod_{l=1}^{\nu-1} \sin^2 \alpha_{\mu l} + 2 \sin \alpha_{\mu \nu} \cos \alpha_{\mu l} \prod_{l=1, l \neq \nu}^{r-1} \sin^2 \alpha_{\mu l} \\
& + \sum_{p=\nu+1}^{r-1} 2 \cos \alpha_{\mu p} \sin \alpha_{\mu \nu} \cos \alpha_{\mu \nu} \prod_{l=1, l=\mu}^{p-1} \sin^2 \alpha_{\mu l}),
\end{aligned}$$

从而利用 Armijo 线搜索下的非线性共轭梯度方法求解问题 (2.2) 的算法如下.

算法 2.1 (Armijo 线搜索下的非线性共轭梯度方法求解问题 (2.2))

步 0 给定参数 $\rho \in (0, 1), \delta \in (0, 0.5), \sigma \in (\delta, 0.5)$, 迭代精度 $0 \leq \varepsilon < 1$ 和初始点 α_0 . 计算 $g_0 = \nabla F(\alpha_0)$. 令 $k := 0$.

步 1 若 $\|g_k\| \leq \varepsilon$, 停机, 输出 $\alpha^* \approx \alpha_k$.

步 2 计算搜索方向

$$d_k = \begin{cases} -g_k, k = 0, \\ -g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} d_{k-1}, k > 0. \end{cases}$$

步 3 利用 Armijo 法则确定搜索步长 b_k 即找一个最小的非负整数 m_k , 使得 $F(\alpha_{k+1}) \leq F(\alpha_k) + \sigma \rho^{m_k} g_k^T d_k$. 令 $b_k = \rho^{m_k}$, $\alpha_{k+1} = \alpha_k + b_k d_k$.

步 4 令 $k := k + 1$, 转步 1.

由文 [12] 的定理 2.6 得到算法 2.1 的全局收敛性定理.

定理 2.2 设问题 (2.2) 目标函数 $F(\alpha)$ 连续可微, 如果问题 (2.2) 的梯度是 Lipschitz 连续的, 即存在常数 $L > 0$, 使得 $\|\nabla F(\alpha_1) - \nabla F(\alpha_2)\|_F \leq L \|\alpha_1 - \alpha_2\|_F, \forall \alpha_1, \alpha_2 \in \mathbf{R}^{m \times (r-1)}$, 那么由算法 2.1 产生的序列 $\{\alpha_k\}$ 满足 $\liminf_{k \rightarrow \infty} \|\nabla F(\alpha_k)\|_F = 0$.

3 数值例子

本节利用数值例子验证算法 2.1 求解问题 (2.2) 是可行的. 所有程序都用 Matlab R2014a 编写. 令梯度范数为 $\|\nabla F_k\|_F = \|\nabla F(\alpha_k)\|_F$, 停机标准为 $\|\nabla F_k\|_F < 1 \times 10^{-3}$, 其中 α_k 表示算法 2.1 的第 k 步迭代值.

例 3.1 考虑问题 (2.2), 随机选取加权无向图 (见图 1), 其邻接矩阵为

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

从而图 1 的负 Laplace 矩阵为

$$Q = \frac{1}{4}(\text{diag}(Ae) - A) = \begin{bmatrix} 0.50 & -0.25 & 0 & 0 & -0.25 & 0 \\ -0.25 & 0.75 & -0.25 & 0 & -0.25 & 0 \\ 0 & -0.25 & 0.50 & -0.25 & 0 & 0 \\ 0 & 0 & -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0 & -0.25 & 0.75 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0.25 \end{bmatrix},$$

选取初值

$$\alpha_0 = \begin{bmatrix} 0.0594 & 0.0924 & 0.1088 \\ 0.3158 & 0.0078 & 0.6318 \\ 0.7727 & 0.4231 & 0.1265 \\ 0.6964 & 0.6556 & 0.1343 \\ 0.1253 & 0.7229 & 0.0986 \\ 0.1302 & 0.5312 & 0.1420 \end{bmatrix}.$$

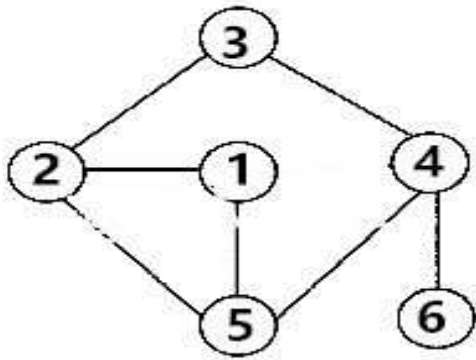
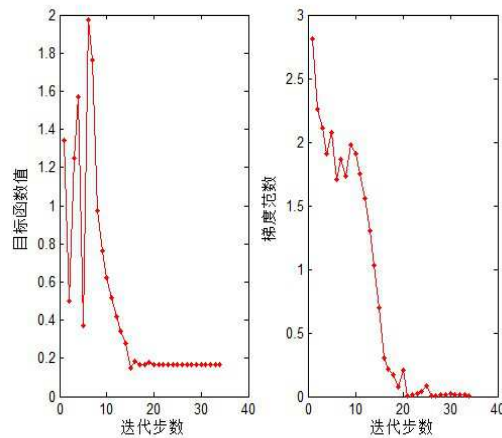


图 1: 加权无向图 (各点权重相同)

图 2: 目标函数值和梯度范数 $\|\nabla F_k\|_F$ 的曲线

利用算法 2.1 求解问题 (2.2). 其目标函数值和梯度范数 $\|\nabla F_k\|_F$ 的曲线如图 2.

例 3.2 考虑问题 (2.2), 随机选取一幅加权无向图并按例子 3.1 所述方法求得其负 Laplace 矩阵为 Q , 再利用算法 2.1 求解, 其中

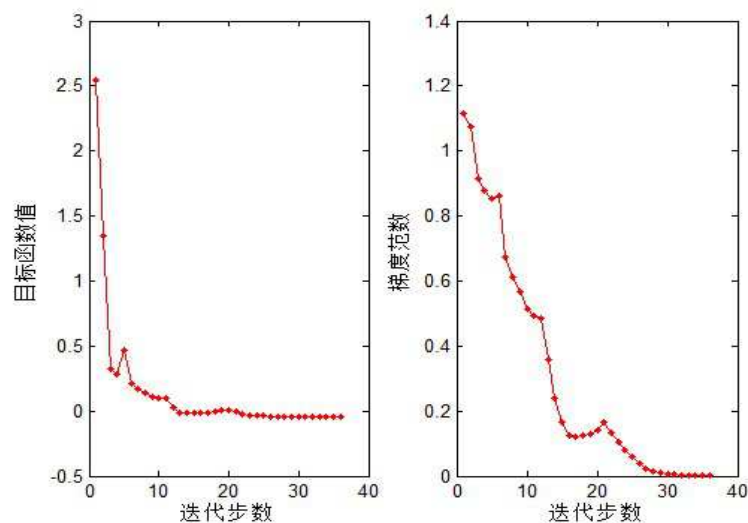
$$Q = \begin{bmatrix} 0.3800 & 0.0110 & 0.0071 & 0.0010 & 0.0020 & 0.0060 & 0.0011 & 0.0087 & 0.0067 & 0.0980 \\ 0.0110 & 0.0830 & 0.0085 & 0.0100 & 0.0012 & 0.0100 & 0.1000 & 0.0590 & 0.0018 & 0.0064 \\ 0.0460 & 0.0033 & 0.0320 & 0.0100 & 0.0068 & 0.0080 & 0.0115 & 0.0290 & 0.0044 & 0.0240 \\ 0.0900 & 0.0083 & 0.0870 & 0.0480 & 0.0060 & 0.0660 & 0.0077 & 0.0860 & 0.0065 & 0.0740 \\ 0.0520 & 0.0090 & 0.0094 & 0.0340 & 0.0560 & 0.0046 & 0.0071 & 0.0600 & 0.0910 & 0.0099 \\ 0.0490 & 0.0250 & 0.0120 & 0.0056 & 0.0230 & 0.0960 & 0.0095 & 0.0390 & 0.0077 & 0.0024 \\ 0.0360 & 0.0094 & 0.0330 & 0.0100 & 0.0068 & 0.0200 & 0.0690 & 0.0780 & 0.0047 & 0.0060 \\ 0.0930 & 0.0024 & 0.0093 & 0.0059 & 0.0076 & 0.0022 & 0.1300 & 0.0540 & 0.0062 & 0.0084 \\ 0.0450 & 0.0300 & 0.0590 & 0.0078 & 0.0009 & 0.0034 & 0.0006 & 0.0270 & 0.0810 & 0.0051 \\ 0.0700 & 0.0061 & 0.0032 & 0.0011 & 0.0420 & 0.0040 & 0.0740 & 0.0210 & 0.0042 & -0.0770 \end{bmatrix}.$$

选取初值

$$\alpha_0 = \begin{bmatrix} 0.1467 & 0.7523 & 0.8600 & 0.5047 & 0.8818 \\ 0.3513 & 0.0959 & 0.7497 & 0.5120 & 0.9157 \\ 0.3623 & 0.3418 & 0.3150 & 0.8594 & 0.7320 \\ 0.2052 & 0.3213 & 0.5096 & 0.0467 & 0.8598 \\ 0.5802 & 0.3489 & 0.7831 & 0.7015 & 0.8034 \\ 0.5385 & 0.4917 & 0.3721 & 0.8031 & 0.2727 \\ 0.7920 & 0.5877 & 0.7897 & 0.1076 & 0.6315 \\ 0.2900 & 0.2038 & 0.2084 & 0.8616 & 0.0347 \\ 0.9569 & 0.2725 & 0.1200 & 0.4101 & 0.1399 \\ 0.0800 & 0.2397 & 0.3625 & 0.5672 & 0.0335 \end{bmatrix},$$

利用算法 2.1 求解问题 (2.2), 可得其目标函数值和梯度范数的曲线如下图 3.

例 3.3 考虑问题 (2.2), 其中 $Q \in \mathbf{R}^{n \times n}$, $\alpha_0 \in \mathbf{R}^{n \times (r-1)}$ 随机选取, 且 $r \leq n$. 利用算法 2.1 求解问题 (2.2), 实验结果如表 1, 其中 IT 表示迭代步数, CPU 表示 CPU 运行时间, VAL

图 3: 目标函数值和梯度范数 $\|\nabla F_k\|_F$ 的曲线

表示目标函数值, GN 表示梯度范数.

表 1: n 和 r 取不同值时算法 3.2 的结果

n, r	3, 2	5, 4	10, 6	15, 8	20, 11
IT	12	28	36	194	312
CPU(S)	0.004 9	0.039 2	0.259 3	5.440 0	21.840 0
VAL	-0.267 3	-0.021 3	-0.041 1	-0.108 1	-0.012 0
GN	0.001 0	0.001 0	0.001 0	0.001 0	0.001 0

数值例子 3.1、3.2 和 3.3 说明利用算法 2.1 求解问题 (2.2) 是可行的.

4 结论

本文考虑图像处理中的最小割问题, 利用 Gramian 表示和三角函数变换将图像处理中的最小割问题转化为无约束优化问题, 再利用非线性共轭梯度法求解无约束优化问题, 最后用数值实验验证了迭代方法是可行的.

参 考 文 献

- [1] Grippo L, Palagi L, Piccialli V. An unconstrained minimization method for solving low-rank SDP relaxations of the maxcut problem[J]. Mathematical Programming (Series B), 2011, 126(1): 119–146.
- [2] 罗希平, 田捷, 诸葛婴, 等. 图像分割方法综述 [J]. 模式识别与人工智能, 1999, 12(3): 300–312.
- [3] Kokiopoulou E, Chen J, Saad Y. Trace optimization and eigenproblems in dimension reduction methods[J]. Numerical Linear Algebra with Applications, 2011, 18(3): 565–602.
- [4] Weinberger K. Q., Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10(1): 207–244.

- [5] Kiers H A L, Berge J M F T. Minimization of a class of matrix trace functions by means of refined majorization[J]. Psychometrika, 1992, 57(3): 371–382.
- [6] Bental A, Teboulle M. Hidden convexity in some nonconvex quadratically constrained quadratic programming[J]. Mathematical Programming, 1996, 72(1): 51–63.
- [7] Benson S J, Ye Y, Zhang X. Solving large-scale sparse semidefinite programs for combinatorial optimization[J]. SIAM Journal on Optimization, 2000, 10(2): 443–461.
- [8] Burer S, Monteiro R D C, et al. Rank-two relaxation Heuristics for MAX-CUT and other binary quadratic programs[J]. SIAM Journal on Optimization, 2006, 12(2): 503–521.
- [9] Liu X G, Wang X F, Wang W G. Maximization of matrix trace function of product stiefel manifolds[J]. SIAM Journal on Matrix Analysis and Applications, 2015, 36(4): 1489–1506.
- [10] Xu S, Gao L, Zhang P. Numerical linear algebra[M], Beijing: Peking University Press, 2010.
- [11] Duan X F, Bai J C, Zhang M J. On the generalized low rank approximation of the correlation matrices arising in the asset portfolio[J]. Linear Algebra and Its Applications, 2014, 461: 1–17.
- [12] Adrian S, Lewis, Malick J. Alternating projections on manifolds[J]. Mathematics of Operations Research, 2008, 33(1): 216–234.

THE NONLINEAR CONJUGATE GRADIENT METHOD FOR SOLVING A CLASS OF THE MATRIX TRACE MINIMIZATION PROBLEM

LI Chun-mei^{1,2}, WANG Cui-fang², DUAN Xue-feng²

(1.School of Mathematical Science, Guizhou Normal University, Guiyang 550001, China)

(2.School of Mathematics and Computational Science, Guilin University of Electronic Technology,
Guilin 541004, China)

Abstract: In this paper, we consider the trace minimization problem in graph partitioning. Using the Gramian representation of the positive semidefinite matrix, the problem can be formulated as the unconstrained optimization problem, then the nonlinear conjugate method with the Armijo line search is used to solve it. Numerical experiments illustrate the feasibility of the new method.

Keywords: matrix trace minimization; Gramian representation; nonlinear conjugate gradient method

2010 MR Subject Classification: 15A15; 49M99