



非负矩阵低秩分解的交替二次规划算法

阳明盛^{*1}, 刘力军²

(1. 大连理工大学 数学科学学院, 辽宁 大连 116024;
2. 大连民族学院 理学院, 辽宁 大连 116600)

摘要: 非负矩阵分解算法有多种, 但都存在着各自的缺陷. 在现有工作的基础上, 将非负矩阵分解(NMF)模型转化为一组(两个)二次凸规划模型, 利用二次凸规划有解的充分必要条件推导出迭代公式, 进行交替迭代, 可求出问题的解. 得到的解不仅具有某种最优性、稀疏性, 还避免了约束非线性规划求解的复杂过程和大量的计算. 证明了迭代的收敛性, 且收敛速度快于已知的方法, 对于大规模数据模型尤能显示出其优越性.

关键词: 非负矩阵分解; 二次凸规划; 大规模数据模型

中图分类号: O224

文献标识码: A

doi: 10.7511/dllgxb201403016

0 引言

在机器学习和模式识别的应用^[1-3]中, 通过矩阵分解得到原数据矩阵的低秩逼近, 可以发现数据的内在结构特征, 从而大大降低数据特征的维数, 节省存储和计算资源. Lee等^[4]发表了非负矩阵分解(non-negative matrix factorization, NMF)算法, 它将一个元素非负的矩阵分解为左右两个非负矩阵乘积, 由于分解后的矩阵中仅包含非负元素, 原矩阵中列向量可解释为对左矩阵中所有列向量(称为基向量)的加权和, 而权重系数为右矩阵中对应列向量中的元素. 由于其分解算法实现简便, 分解的结果中不出现负值, 而且具有可解释性和明确的物理意义, 广泛应用于图像识别、语音辨识、盲源分离、脑电信号提取等问题中^[3-6]. 本文将 NMF 模型转化为凸二次规划模型求解, 以期提高大规模数据模型的处理速度.

1 非负矩阵分解的概念与计算方法

给定 $m \times n$ 阶的非负实矩阵 A 和正整数 $k < \min\{m, n\}$, NMF 要求找到两个非负矩阵 $W \in R^{m \times k}$ 和 $H \in R^{k \times n}$, 满足

$$A \approx WH \quad (1)$$

衡量 WH 与 A 之间逼近程度误差的度量有很多, 采用不同的度量可以得到不同的优化模型.

基于 Frobenius 范数的欧氏距离优化模型^[4-5]如下:

$$\min_{W \geq 0, H \geq 0} E(A \| WH) = \frac{1}{2} \|A - WH\|_F^2 \quad (2)$$

或者

$$\min_{W \geq 0, H \geq 0} E(A \| WH) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (A - WH)_{ij}^2 \quad (3)$$

若选用 K-L 散度函数作为逼近度量, 则问题(1)可以转化为下面的优化模型^[4, 6-7]:

$$\min_{W \geq 0, H \geq 0} E(A \| WH) = \sum_{i=1}^m \sum_{j=1}^n \left(A_{ij} \log \frac{A_{ij}}{(WH)_{ij}} - A_{ij} + (WH)_{ij} \right) \quad (4)$$

非负矩阵低秩分解的方法有很多^[6-11], 思路可分两类. 首先是公式逼近法, 根据式(1)的要求, 构造某种迭代公式, 其中最基本最具代表性的是^[4-5, 7]

$$\begin{aligned} H_{k+1} &= H_k * (W_k^T A) / (W_k^T W_k H_k) \\ W_{k+1} &= W_k * (A H_k^T) / (W_k H_k H_k^T) \end{aligned} \quad (5)$$

采用同时迭代或者交替迭代的方式, 将 H_k 与 W_k 更新为 H_{k+1} 与 W_{k+1} , 并希望当 $k \rightarrow +\infty$ 时, $W_k H_k \rightarrow A$.

公式逼近法的优点是, 迭代公式的构造一般

都比较简单,公式中的运算主要是矩阵之间加法、乘法与矩阵求逆运算等,算法的实现比较容易;缺点是其收敛性难以得到证明,且收敛速度较慢。

其次是优化逼近法。针对模型(3)直接采用优化方法求解,可供选择的局部最优化方法^[12-16]有广义乘子法、广义既约梯度法(GRG法)、信赖域方法、二次规划序列逼近法、梯度投影法等。其主要优点是算法都比较成熟,可靠性高,收敛性都已得到充分证明;其缺点是求得的解只能保证它的局部最优性。亦可选择全局最优化方法求解^[15-16],如填充函数法、山丘函数法、区间方法、遗传算法、微分进化算法、随机投点与随机方向算法等。这类方法的缺点主要是计算量远远大于一般的局部最优化方法。

2 非负矩阵低秩分解的交替二次规划算法

在对模型(3)采用局部或全局优化计算时,可看到 $f(\mathbf{X})$ 是一个多元($p = m \times k + k \times n$ 个变元)四次代数多项式函数,由于其高度非线性特性,极值点结构非常复杂,使用传统的优化算法求解模型(3)十分困难。注意到在模型(3)中,如果将 $\mathbf{W}$ 选定为某个常数阵 $\mathbf{W}_1$,即令 $\mathbf{W} = \mathbf{W}_1$ 时有

$$\min_{\mathbf{H} \geqslant \mathbf{0}} E(\mathbf{A} \parallel \mathbf{H}) = \min_{\mathbf{H} \geqslant \mathbf{0}} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W}_1 \mathbf{H})_{ij}^2 \tag{6}$$

是一个关于 $\mathbf{H}$ 的二次函数,而约束 $\mathbf{H} \geqslant \mathbf{0}$ 是关于 $\mathbf{H}$ 的一次函数,因而式(6)是一个关于 $\mathbf{H}$ 的二次规划模型。

同理,如果选定 $\mathbf{H} = \mathbf{H}_1$,则有

$$\min_{\mathbf{W} \geqslant \mathbf{0}} E(\mathbf{A} \parallel \mathbf{W}) = \min_{\mathbf{W} \geqslant \mathbf{0}} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W} \mathbf{H}_1)_{ij}^2 \tag{7}$$

是一个关于 $\mathbf{W}$ 的二次规划模型。如果在式(6)与(7)中再加入相应的稀疏性约束 $\lambda \sum_{i=1}^m \sum_{j=1}^n |H_{ij}|$ 与 $\mu \sum_{i=1}^m \sum_{j=1}^k |W_{ij}|$,有

$$\min_{\mathbf{H} \geqslant \mathbf{0}} E(\mathbf{A} \parallel \mathbf{H}) = \min_{\mathbf{H} \geqslant \mathbf{0}} \left[\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W}_1 \mathbf{H})_{ij}^2 + \lambda \sum_{i=1}^k \sum_{j=1}^n |H_{ij}| \right]; \lambda > 0$$

$$\min_{\mathbf{W} \geqslant \mathbf{0}} E(\mathbf{A} \parallel \mathbf{W}) = \min_{\mathbf{W} \geqslant \mathbf{0}} \left[\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W} \mathbf{H}_1)_{ij}^2 + \mu \sum_{i=1}^m \sum_{j=1}^k |W_{ij}| \right]; \mu > 0$$

$$\mu \sum_{i=1}^m \sum_{j=1}^k |W_{ij}| \Big]; \mu > 0$$

式中: $|H_{ij}|$ 为矩阵 $\mathbf{H}$ 任一元素的绝对值, $|W_{ij}|$ 为矩阵 $\mathbf{W}$ 任一元素的绝对值, λ 与 μ 为乘子或惩罚因子,以上两式又可写成:

$$\min_{\mathbf{H} \geqslant \mathbf{0}} E(\mathbf{A} \parallel \mathbf{H}) = \min_{\mathbf{H} \geqslant \mathbf{0}} \left[\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W}_1 \mathbf{H})_{ij}^2 + \lambda \sum_{i=1}^k \sum_{j=1}^n H_{ij} \right]; \lambda > 0 \tag{8}$$

$$\min_{\mathbf{W} \geqslant \mathbf{0}} E(\mathbf{A} \parallel \mathbf{W}) = \min_{\mathbf{W} \geqslant \mathbf{0}} \left[\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W} \mathbf{H}_1)_{ij}^2 + \mu \sum_{i=1}^m \sum_{j=1}^k W_{ij} \right]; \mu > 0 \tag{9}$$

显然式(8)与(9)均是二次凸规划模型,当乘子(或参数) λ 与 μ 选定以后,它们的解 $\mathbf{H}$ 与 $\mathbf{W}$ 是存在的、唯一的。

求解二次凸规划模型的难度远比求解模型(3)要小得多。求解方法有 Lemke 算法、Wolfe 算法、有效集算法等^[12-15],其中有效集算法对二次规划中二次项系数阵的要求比较宽松,一般的数学软件优化计算部分中,对二次规划的求解大多采用有效集算法,例如 Matlab 7.0 软件优化库程序中对二次规划的求解就是如此^[17]。

如果充分利用函数 $E(\mathbf{A} \parallel \mathbf{H})$ 与 $E(\mathbf{A} \parallel \mathbf{W})$ 的凸性,还可以有更简捷的办法来进行求解,式(8)与(9)中约束极小化的目标函数又可写成

$$E(\mathbf{A} \parallel \mathbf{H}) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W}_1 \mathbf{H})_{ij}^2 + \lambda \sum_{i=1}^k \sum_{j=1}^n H_{ij}; H_{ij} \geqslant 0, \lambda > 0 \tag{10}$$

与

$$E(\mathbf{A} \parallel \mathbf{W}) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W} \mathbf{H}_1)_{ij}^2 + \mu \sum_{i=1}^m \sum_{j=1}^k W_{ij}; W_{ij} \geqslant 0, \mu > 0 \tag{11}$$

求出 $\frac{\partial E(\mathbf{A} \parallel \mathbf{H})}{\partial \mathbf{H}}$ 与 $\frac{\partial E(\mathbf{A} \parallel \mathbf{W})}{\partial \mathbf{W}}$,并令 $\frac{\partial E(\mathbf{A} \parallel \mathbf{H})}{\partial \mathbf{H}} = \mathbf{0}$ 与 $\frac{\partial E(\mathbf{A} \parallel \mathbf{W})}{\partial \mathbf{W}} = \mathbf{0}$,这是式(10)与(11)有解(有极值点)的必要条件,同时也是充分性条件(因为是凸规划),容易求得

$$\frac{\partial E(\mathbf{A} \parallel \mathbf{H})}{\partial H_{ij}} = (\mathbf{W}_1^T (\mathbf{W}_1 \mathbf{H} - \mathbf{A}))_{ij} + \lambda = (\mathbf{W}_1^T \mathbf{W}_1 \mathbf{H} - \mathbf{W}_1^T \mathbf{A})_{ij} + \lambda$$

$$\text{令 } \frac{\partial E(\mathbf{A} \parallel \mathbf{H})}{\partial \mathbf{H}} = \mathbf{0}, \text{有}$$

$$\mathbf{W}_1^T \mathbf{W}_1 \mathbf{H} = \mathbf{W}_1^T \mathbf{A} - \lambda \mathbf{L}$$

式中： $\mathbf{L}$ 为 $k \times n$ 阶、元素全为1的矩阵。从而可得

$$\mathbf{H} = (\mathbf{W}_1^T \mathbf{W}_1)^+ (\mathbf{W}_1^T \mathbf{A} - \lambda \mathbf{L})$$

其中 $\mathbf{A}$ 为 $m \times n$ 阶矩阵， $\mathbf{W}_1$ 为 $m \times k$ 阶矩阵， $(\mathbf{W}_1^T \mathbf{W}_1)^+$ 是 $\mathbf{W}_1^T \mathbf{W}_1$ 的Penrose-Moore广义逆阵^[17]，同理可以求得

$$\frac{\partial E(\mathbf{A} \parallel \mathbf{W})}{\partial \mathbf{W}_{ij}} = ((\mathbf{W} \mathbf{H}_1 - \mathbf{A}) \mathbf{H}_1^T)_{ij} + \mu =$$

$$(\mathbf{W} \mathbf{H}_1 \mathbf{H}_1^T - \mathbf{A} \mathbf{H}_1^T)_{ij} + \mu$$

$$\text{令 } \frac{\partial E(\mathbf{A} \parallel \mathbf{W})}{\partial \mathbf{W}} = \mathbf{0}, \text{ 有 } \mathbf{W} \mathbf{H}_1 \mathbf{H}_1^T = \mathbf{A} \mathbf{H}_1^T - \mu \mathbf{M},$$

其中 $\mathbf{M}$ 为 $m \times k$ 阶、元素全为1的矩阵，从而得

$$\mathbf{W} = (\mathbf{A} \mathbf{H}_1^T - \mu \mathbf{M})(\mathbf{H}_1 \mathbf{H}_1^T)^+$$

其中 $\mathbf{H}_1$ 为 $k \times n$ 阶矩阵， $(\mathbf{H}_1 \mathbf{H}_1^T)^+$ 是 $\mathbf{H}_1 \mathbf{H}_1^T$ 的Penrose-Moore广义逆阵^[17]。

则交替二次规划算法的计算公式可以归纳如下：

$$\mathbf{H} = (\mathbf{W}_1^T \mathbf{W}_1)^+ (\mathbf{W}_1^T \mathbf{A} - \lambda \mathbf{L}); \lambda > 0, \mathbf{H} \geq \mathbf{0} \quad (12)$$

$$\mathbf{W} = (\mathbf{A} \mathbf{H}_1^T - \mu \mathbf{M})(\mathbf{H}_1 \mathbf{H}_1^T)^+; \mu > 0, \mathbf{W} \geq \mathbf{0} \quad (13)$$

$$E_{2k-1} = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W}_1 \mathbf{H})_{ij}^2 + \lambda \sum_{i=1}^k \sum_{j=1}^n H_{ij} = E_a + E_b \quad (14)$$

$$E_{2k} = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{A} - \mathbf{W} \mathbf{H}_1)_{ij}^2 + \mu \sum_{i=1}^m \sum_{j=1}^k W_{ij} = E_a + E_b \quad (15)$$

具体计算步骤为

步骤1 选取基矩阵 $\mathbf{W}$ 的初值为 $\mathbf{W} = \mathbf{W}_k (k = 1)$ ，按式(12)求 $\mathbf{H}_k (k = 1)$ ，注意其中参数 λ 试探性地取值，以使得到的 $\mathbf{H}_k (k = 1)$ 中正元素的多少与分布都比较合理，然后对 $\mathbf{H}_k$ 引入非负性条件限制，记录下 $\mathbf{W}_k$ 与 $\mathbf{H}_k (k = 1)$ 中零元素的个数(设为 $N_{\mathbf{W}_k z}$ 与 $N_{\mathbf{H}_k z}$)，再按式(14)计算总误差 $E_{2k-1} (k = 1)$ ；

步骤2 取定 $\mathbf{H} = \mathbf{H}_k$ ，按式(13)求 $\mathbf{W}_{k+1}$ ，得到 $\mathbf{W}_{k+1}$ 后引入非负性条件限制，注意其中参数 μ 试探性地取值，以使得到的 $N_{\mathbf{W}_{k+1} z}$ 满足 $N_{\mathbf{W}_{k+1} z} \geq N_{\mathbf{W}_k z}$ ，再按式(15)计算总误差 $E_{2k} (k = 1)$ ；

步骤3 检查迭代终止条件：如果 $E_{2k-1} \geq E_{2k} (k = 1)$ 或者 $E_{2k} \geq E_{2k+1}$ 与 $N_{\mathbf{W}_2 z} \geq N_{\mathbf{W}_1 z}$ 同时成立，则令 $k + 1 \rightarrow k$ ，转入步骤1继续迭代，否则迭代终止，上一轮的结果则为最优分解。

注：1. 初始基阵 $\mathbf{W}_k (k = 1)$ 的选取方法：只需 $\mathbf{W}_1$ 中的列向量尽可能地线性独立，并含有尽可能多的零元素即可。初始基阵 $\mathbf{W}_1$ 选取得好与不好，

均不影响交替二次规划算法迭代的收敛性，但 $\mathbf{W}_1$ 选取得好可以减少迭代的次数。

2. 参数 λ 与 μ 的取值方法：只需满足两个条件：(1) 从 $\mathbf{H}_k$ 到 $\mathbf{H}_{k+1}$ 中零元素的个数 $N_{\mathbf{H}_k z}$ 到 $N_{\mathbf{H}_{k+1} z}$ 是缓慢递增的，即有 $N_{\mathbf{H}_k z} \leq N_{\mathbf{H}_{k+1} z}$ (或从 $\mathbf{W}_k$ 到 $\mathbf{W}_{k+1}$ 中零元素的个数 $N_{\mathbf{W}_k z}$ 到 $N_{\mathbf{W}_{k+1} z}$ 是缓慢递增的)；(2) 按照式(14)或(15)计算出的前后两次总误差值 E_{2k-1} 到 E_{2k} 或 E_{2k} 到 E_{2k+1} 是逐渐递减的，即有 $E_{2k-1} \geq E_{2k} \geq E_{2k+1}$ 。如果所取 λ 或 μ 使条件(1)、(2)不能同时满足，则迭代终止；如果能够同时满足则继续向下迭代。在具体执行中，比如下面例1，当求 $\mathbf{H}_2$ 时，如果 λ 取值过大，若取 $\lambda = 16$ ，则求得的 $\mathbf{H}_2$ 中，或者全部元素为零，或者仅有少量非负元素；如果 λ 取值过小，例如取 $\lambda = 1$ ，则求得的 $\mathbf{H}_2$ 中，或者没有零元素，或者仅有少量的零元素。需要对 λ 值进行调整，如过大，只需将 $\lambda = 16$ 除以2有 $\lambda = 16/2 = 8$ ，如过小只需将 $\lambda = 1$ 乘以2有 $\lambda = 1 \times 2 = 2$ ，直到 λ 值使 $N_{\mathbf{H}_2 z} \geq N_{\mathbf{H}_1 z}$ 为止。

3. $\mathbf{H}_k$ 与 $\mathbf{W}_k$ 中元素非负性要求的实现：当模型取为一般的约束非线性规划(3)时，则 $\mathbf{H} \geq \mathbf{0}$ ， $\mathbf{W} \geq \mathbf{0}$ ，将构成模型(3)可行域的(线性)边界，通常采用梯度投影法效果较好，但因需作多次的一维搜索，计算量往往较大。当模型取为交替二次规划(10)、(11)时，多数文章采用二次规划(10)或(11)有解的必要性条件(Kuhn-Tucker条件)，将式(10)或(11)转化为一组等式和不等式方程联立求解，其中因为存在有非线性的互补松弛性等式，也给求解带来较大的难度与较大的计算量。为了避免上述方法中存在的这些问题，本文在交替二次规划迭代计算中采用当 $\mathbf{H}_k$ 与 $\mathbf{W}_k$ 中出现负元素时，直接用零元素取代，而以较小的难度与较少的计算量实现了 $\mathbf{H}_k$ 与 $\mathbf{W}_k$ 中元素的非负性要求。

3 交替二次规划算法实例

例1 设有非负实矩阵 $\mathbf{A}_{5 \times 5}$ ，如果取分解参数 $r = 3$ ，则有 $\mathbf{A}_{5 \times 5} \approx \mathbf{W}_{5 \times 3} \times \mathbf{H}_{3 \times 5}$ 。

注：当 $r = 1, 2, 4, 5$ 时，求解思路与演算过程完全同于 $r = 3$ 。

当已给 $\mathbf{A}$ ，如果将基矩阵 $\mathbf{W}$ 的初值选定为 $\mathbf{W}_1$ ，则按交替二次规划算法计算式(12)、(13)与(14)、(15)来交替求解 $\mathbf{H}$ 与 $\mathbf{W}$ ，计算过程与具体步骤略去，可依次得分解的主要步结果如下：

(1) $\mathbf{A} = \begin{pmatrix} 11 & 1 & 1 & 1 & 1 \\ 1 & 21 & 1 & 1 & 1 \\ 1 & 1 & 31 & 1 & 1 \\ 1 & 1 & 1 & 41 & 1 \\ 1 & 1 & 1 & 1 & 51 \end{pmatrix},$

$\mathbf{W}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$

$\mathbf{H}_1 = \begin{pmatrix} 2 & 0 & 12 & 0 & 0 \\ 0 & 7 & 0 & 17 & 0 \\ 0 & 0 & 0 & 0 & 43 \end{pmatrix},$

$E_1 = 1\,502, \lambda = 8;$

(2) $\mathbf{W}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.130\,17 & 0 \\ 1.716\,22 & 0 & 0 \\ 0 & 1.727\,81 & 0 \\ 0 & 0 & 1.121\,15 \end{pmatrix},$

$\mathbf{H}_1$ 不变, $E_2 = 1\,025.1, \mu = 120;$

(3) $\mathbf{W}_2$ 不变,

$\mathbf{H}_2 = \begin{pmatrix} 0 & 0 & 16.704\,9 & 0 & 0 \\ 0.153\,7 & 0 & 0 & 22.306\,6 & 0 \\ 0 & 0 & 0 & 0 & 42.306\,8 \end{pmatrix},$

$E_3 = 630.053, \lambda = 4;$

(4) $\mathbf{W}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1.545\,01 & 0 & 0 \\ 0 & 1.681\,07 & 0 \\ 0 & 0 & 1.192\,27 \end{pmatrix},$

$\mathbf{H}_2$ 不变, $E_4 = 1\,025.1, \mu = 100.$

迭代至第(4)步停止,因为无论怎样调整 $\mu > 0$ 值,均不能使 $E_4 \leqslant E_3$ 与 $N_{\mathbf{W}_3z} \geqslant N_{\mathbf{W}_2z}$ 同时成立,故最优分解是第(3)步的结果 $\mathbf{W}_2, \mathbf{H}_2$.

例 2 设有非负实矩阵 $\mathbf{A}_{7 \times 6}$,如果取分解参数 $k = 3$,则有 $\mathbf{A}_{7 \times 6} \approx \mathbf{W}_{7 \times 3} \times \mathbf{H}_{3 \times 6}$.

注:当 $k = 1, 2, 4, 5, 6$ 时,求解思路与演算过程完全同于 $k = 3$.

已给 $\mathbf{A}$,如果将基矩阵 $\mathbf{W}$ 的初值选定为 $\mathbf{W}_1$,则按交替二次规划法求解式(12)、(13)、(14)、(15)来交替求解 $\mathbf{H}$ 与 $\mathbf{W}$,计算过程与具体步骤略去,可依次得分解的主要步结果如下:

(1) $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \\ 3 & 4 & 5 & 6 & 1 & 2 \\ 4 & 5 & 6 & 1 & 2 & 3 \\ 5 & 6 & 1 & 2 & 3 & 4 \\ 6 & 1 & 2 & 3 & 4 & 5 \\ 7 & 5 & 4 & 3 & 2 & 1 \end{pmatrix},$

$\mathbf{W}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$

$\mathbf{H}_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix},$

$E_1 = 319, \lambda = 7;$

(2) $\mathbf{A}$ 不变, $\mathbf{W}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1.5 \end{pmatrix},$

$\mathbf{H}_1$ 不变, $E_2 = 306.25, \mu = 7;$

(3) $\mathbf{W}_2$ 不变,

$\mathbf{H}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 2.941\,18 & 1.294\,12 & 1.411\,76 & 1.529\,41 & 0.235\,29 & 0.352\,94 \end{pmatrix},$

$E_3 = 289.471, \lambda = 7;$

(4) $\mathbf{W}_3 = \begin{pmatrix} 0.515\,00 & 0 & 0.832\,99 \\ 0 & 0.130\,14 & 1.464\,43 \\ 0 & 0 & 2.095\,85 \\ 0 & 0 & 1.707\,28 \\ 0 & 0 & 1.814\,14 \\ 0 & 0 & 1.921\,00 \\ 0 & 0 & 2.688\,43 \end{pmatrix},$

$\mathbf{H}_2$ 不变, $E_4 = 198.176, \mu = 6.5;$

(5) $\mathbf{W}_3$ 不变,

$\mathbf{H}_3 = \begin{pmatrix} 0 & 0 & 0 & 0.718\,41 & 3.240\,88 & 4.818\,04 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 2.993\,17 & 2.557\,94 & 2.277\,46 & 2.026\,77 & 1.667\,73 & 1.893\,13 \end{pmatrix},$

$E_5 = 100.885, \lambda = 1;$

$$(6) \mathbf{W}_4 = \begin{pmatrix} 0.626\ 84 & 0 & 0.408\ 43 \\ 0 & 0 & 0.891\ 79 \\ 0 & 0 & 1.245\ 81 \\ 0 & 0 & 1.210\ 79 \\ 0 & 0 & 1.027\ 41 \\ 0.200\ 23 & 0 & 0.773\ 01 \\ 0 & 0 & 1.570\ 27 \end{pmatrix},$$

$\mathbf{H}_3$ 不变, $E_6 = 521.563, \mu = 20.$

无论怎样调整 $\mu > 0$ 值,均不能使 $E_6 \leq E_5$ 与 $N_{W_4z} \geq N_{W_3z}$ 同时成立,迭代至第(6)步停止,从而第(5)步的结果 $\mathbf{W}_3$ 、 $\mathbf{H}_3$ 是寻求的最佳结果.

4 交替二次规划算法的收敛性

在式(14)中,因 $E_a > 0, \lambda > 0, H_{ij} \geq 0,$
 $\sum_{i=1}^k \sum_{j=1}^n H_{ij} > 0, E_b > 0,$ 故 $E_{2k-1} > 0.$
在式(15)中,因 $E_a > 0, \mu > 0, W_{ij} \geq 0,$
 $\sum_{i=1}^m \sum_{j=1}^k W_{ij} > 0, E_b > 0,$ 故 $E_{2k} > 0.$

由交替二次规划所产生的总误差值列有 $E_1 \geq E_2 \geq E_3 \geq \cdots \geq E_p \geq 0$,单调递减且有下界的数列 $\{E_p\}(p = 1, 2, 3, \cdots)$ 必有极限,故算法是收敛的,实际上算法在(很少的)有限步终止,其中每一步,如例2中的第(2)、(3)、(4)步的结果,均可作为需要分解的结果,这只需根据 $\mathbf{WH}$ 逼近 $\mathbf{A}$ 的精度 E 值的不同要求来选定.

5 结 语

非负矩阵低秩分解的方法有很多,对于同一个例子用这些方法求得的结果,如何判定(或评定,或比较)它们的优缺点?本文作者认为这需要有一定的评定标准,而这个评定标准中主要应该包括(1)总误差 $E = E_a + E_b$ 的大小,其中 E_a 为 $\mathbf{W}_{m \times k} \times \mathbf{H}_{k \times n}$ 逼近 $\mathbf{A}_{m \times n}$ 所产生的误差, E_b 为满足约束情况所产生的误差,总误差 E 值愈小愈好;(2)基矩阵 $\mathbf{W}$ 中列向量 $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \cdots$,线性独立的程度,这种程度是用内积 $\mathbf{W}_i \times \mathbf{W}_j \approx \mathbf{0}(i \neq j)$ 来描述的, $\mathbf{W}_i \times \mathbf{W}_j$ 愈接近于 $\mathbf{0}$ 愈好;(3)基矩阵 $\mathbf{W}$ 与系数矩阵 $\mathbf{H}$ (特别是 $\mathbf{H}$)稀疏的程度;(4)对大型数据,方法的适应性与处理能力,其中(1)与(3)是统一而又矛盾的.

交替二次规划算法的提出,首先是考虑了避开直接求解原约束非线性规划(3)的复杂性与困难

度.本文在学习与借鉴前人思路的基础上,充分考虑了对评定标准(1)~(4)等基本条件的要求,经过长时间探索而设计出了一种求解方法.实例试算表明,上述思想与要求基本上都能得到很好的体现.算法的思路明确,易于实现,迭代过程简单,计算量很小,对大型数据模型的计算可显示出其特有的优点,算法的收敛性也比较容易得到证明.

缺点是对 $\mathbf{H}_k$ 与 $\mathbf{W}_k$ 中元素非负性要求的实现,采用在迭代过程中如果出现了负元素直接用零元素取代的方案,其优点是比别的任何方法实现过程简单,且计算量很小,但这一方案是否具有普遍性,是否存在不适用的特例,目前尚不清楚,有待进一步的研究.

参考文献:

[1] GAO Yuan, CHURCH G. Improving molecular cancer class discovery through sparse non-negative matrix factorization [J]. *Bioinformatics*, 2005, 21(21):3970-3975.

[2] Hoyer P O. Non-negative matrix factorization with sparseness constraints [J]. *The Journal of Machine Learning Research*, 2004, 5:1457-1469.

[3] “10 000个科学难题”数学编委会. 10 000 个科学难题数学卷 [M]. 北京:科学出版社, 2009.

Mathematics Editorial Board of “10 000 Selected Problems in Sciences”. 10 000 Selected Problems in Sciences-Volume Mathematics [M]. Beijing: Science Press, 2009. (in Chinese)

[4] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization [J]. *Nature*, 1999, 401(6755):788-791.

[5] Lee D D, Seung H S. Algorithms for non-negative matrix factorization [C] // *Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2001.

[6] 蒋娇娇. 非负矩阵分解算法的改进及应用[D]. 北京:北京工业大学, 2011.

JIANG Jiao-jiao. Improvement of non-negative matrix decomposition methods and its applications [D]. Beijing: Beijing University of Technology, 2011. (in Chinese)

[7] 杨洪礼. 非负矩阵与张量分解及其应用[D]. 青岛:山东科技大学, 2011.

YANG Hong-li. Non-negative matrix and tensor decomposition and its applications [D]. Qingdao: Shandong University of Science & Technology, 2011. (in Chinese)

- [8] 张宇飞. 加稀疏约束的非负矩阵分解[D]. 大连:大连理工大学, 2010.
ZHANG Yu-fei. Non-negative matrix decomposition with sparseness constraint [D]. Dalian: Dalian University of Technology, 2010. (in Chinese)
- [9] 赵燕斌. 强噪声中诱发电位信号的快速提取研究[D]. 大连:大连理工大学, 2008.
ZHAO Yan-bin. The fast extraction methods for EP signals with low SNR [D]. Dalian: Dalian University of Technology, 2008. (in Chinese)
- [10] 郭立. 增量式非负矩阵分解方法研究[D]. 上海:复旦大学, 2009.
GUO Li. Research on incremental methods on non-negative matrix decomposition [D]. Shanghai: Fudan University, 2009. (in Chinese)
- [11] 殷海青. 图象分析中的非负矩阵分解理论及其最优化和正则化方法研究[D]. 西安:西安电子科技大学, 2011.
YIN Hai-qing. Theory of non-negative matrix decomposition in image analysis and research on its optimization and regularization methods [D]. Xi'an: Xidian University, 2011. (in Chinese)
- [12] 王德人. 非线性方程组解法与最优化方法[M]. 北京:人民教育出版社, 1979.
WANG Den-ren. **Solutions of Nonlinear Equations and Optimizations Methods** [M]. Beijing: People's Education Press, 1979. (in Chinese)
- [13] 施吉林, 张宏伟, 金光日. 计算机科学计算[M]. 北京:高等教育出版社, 2005.
SHI Ji-lin, ZHANG Hong-wei, JIN Guang-ri. **Computer Scientific Computation** [M]. Beijing: Higher Education Press, 2005. (in Chinese)
- [14] 席少霖, 赵凤治. 最优化计算方法[M]. 上海:上海科学技术出版社, 1983.
XI Shao-lin, ZHAO Feng-zhi. **Optimization Methods** [M]. Shanghai: Shanghai Science and Technology Press, 1983. (in Chinese)
- [15] 张可村, 李换琴. 工程优化方法及其应用[M]. 西安:西安交通大学出版社, 2007.
ZHANG Ke-cun, LI Huan-qin. **Engineering Optimization Methods and Its Applications** [M]. Xi'an: Xi'an Jiaotong University Press, 2007. (in Chinese)
- [16] 阳明盛, 罗长童. 最优化原理、方法及求解软件[M]. 北京:科学出版社, 2006.
YANG Ming-sheng, LUO Chang-tong. **Optimization Principles, Methods and Software** [M]. Beijing: Science Press, 2006. (in Chinese)
- [17] 阳明盛, 熊西文, 林建华. MATLAB 基础及数学软件[M]. 大连:大连理工大学出版社, 2003.
YANG Ming-sheng, XIONG Xi-wen, LIN Jian-hua. **MATLAB Basis and Mathematics Software** [M]. Dalian: Dalian University of Technology Press, 2003. (in Chinese)

Alternative quadratic programming for non-negative matrix low-order factorization

YANG Ming-sheng^{*1}, LIU Li-jun²

(1. School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China;

2. School of Science, Dalian Nationalities University, Dalian 116600, China)

Abstract: Many algorithms are available for solving the problem of non-negative matrix factorization (NMF) despite respective shortcomings. Based on existing works, NMF model is transformed into one group of (two) convex quadratic programming model. Using the sufficient and necessary conditions for quadratic programming problems, iteration formula for NMF is obtained by which the problem is solved after alternative iteration process. The obtained solution reaches its optimality and sparseness while avoiding computational burden and complexity for solving constrained nonlinear programming problems. The iteration convergence can be proved easily and its speed is faster than that of existing approaches. The proposed approach has its superiority for large-scale data model.

Key words: non-negative matrix factorization; convex quadratic programming; large-scale data model