

SENSC: 一个稳定高效的非负稀疏编码算法

李乐^{1,2} 章毓晋^{1,2}

摘要 非负稀疏编码 (Nonnegative sparse coding, NSC) 已成功应用在很多领域的研究中. 目前使用的 NSC 算法通过梯度投影法和基于辅助函数的乘性更新法相结合来实现, 其性能受迭代步长的影响很大, 且效率较低. 为增强 NSC 的可应用性, 本文通过对一组凸超抛物面函数做交替最小化来实现 NSC, 并依据凸超抛物面特性、点到非负数集合的投影规则以及点到原点处单位超球的投影规则构造了一个无用户定义优化参数的稳定高效的 NSC 算法 — SENSC. 从数学角度, 文中推断了 SENSC 比现有算法高效且它的解优于当前算法的解, 证明了它的稳定性和收敛性. 实验验证了上述理论推断的正确, 说明了 SENSC 调节编码稀疏性的能力比已有算法更强.

关键词 非负稀疏编码, 非负矩阵分解, 稀疏编码, 超抛物面
中图分类号 TP391

SENSC: a Stable and Efficient Algorithm for Nonnegative Sparse Coding

LI Le^{1,2} ZHANG Yu-Jin^{1,2}

Abstract Nonnegative sparse coding (NSC) has been successfully applied to many research fields. The applied algorithm for NSC is designed by the combination of gradient projection and auxiliary function-based multiplicative update, so its performance is significantly related to the choice of the iterative step size in gradient projection. Besides, its efficiency can not be very high due to the properties of optimization methods which it uses. To improve the applicability of NSC, we consider the implementation of NSC as alternately minimizing a group of convex hyperparaboloid functions, and propose a stable and efficient NSC algorithm (SENSC) without any user-defined optimization parameter by using the properties of convex hyperparaboloid and the projection formulas from a point to the set of all nonnegative numbers and to the unit super sphere at origin. It is mathematically deduced that SENSC is more efficient than and has solutions superior to the existing algorithm. Its stability and convergence are proven. Experiments have validated theoretical deduction and demonstrated that SENSC is more effective in the control on sparseness of coding results than the existing algorithm.

Key words Nonnegative sparse coding (NSC), nonnegative matrix factorization (NMF), sparse coding (SC), hyper paraboloid

视神经学的一个基本研究目标是探究和了解初级视觉系统的处理过程, 稀疏编码 (Sparse coding, SC) 是进行这项研究的一个有效工具. Olshausen 和 Field 的实验证实对自然图像数据进行 SC 得到的特征可以很好地近似哺乳动物初级视觉突触上简单细胞的感受域^[1]. 由 SC 形成的数据描述与由小波变换 (Wavelet transformation, WT) 形成的数据描述相类似, 但相比 WT, SC 更多地基于数据的统计特性, 因此对数据的自适应性更好^[2].

SC 模型中同时包含“正”和“负”的分解成分, 这与生物学的研究结论相左, 因为神经元的活跃度不会是负数^[2-3], 这也与心理学和生理学的研

究成果不一致, 因为感知是纯加性的生理机能^[4-6]. 因此, Hoyer 将非负矩阵分解 (Nonnegative matrix factorization, NMF)^[7] 的构造思想引入到 SC, 提出了非负稀疏编码 (Nonnegative SC, NSC)^[2-3].

NSC 已成功应用在很多领域的研究中, Hoyer 的实验证实, 不仅可以利用 NSC 有效地模型化哺乳动物初级视觉突触上简单细胞的感受域^[3], 而且对初级视觉突触上复杂细胞因自然图像刺激产生的反应进行 NSC 可以学习出边缘编码 (Contour coding) 和截止区 (End-stopped) 感受域^[8]. 尚丽等的实验表明, 去除图像中的高斯噪声时, NSC 收缩法比 SC 收缩法、WT 收缩法以及维纳滤波法更优^[9]. 此外, Shastri 等构造了基于 NSC 的人脸识别法, 借助 AR 和 YaleB 数据库的实验表明, 此方法可以较好地抑制由表情、光照变化、轻微旋转以及眼部遮挡等带来的不利影响^[10].

NSC 中有两个以相乘形式共存的分解因子, 暂称其为左分解因子和右分解因子. 目前使用的 NSC 算法 — Hoyer 算法, 对左分解因子用梯度投影法优化, 对右分解因子用基于辅助函数的乘性更新法优

收稿日期 2008-04-15 收修改稿日期 2009-01-16
Received April 15, 2008; in revised form January 16, 2009
国家自然科学基金 (60872084) 资助
Supported by National Natural Science Foundation of China (60872084)

1. 清华大学信息科学与技术国家实验室 北京 100084 2. 清华大学电子工程系 北京 100084

1. Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084 2. Department of Electronic Engineering, Tsinghua University, Beijing 100084

DOI: 10.3724/SP.J.1004.2009.01257

化^[2]. 梯度投影法的特点决定其性能受迭代步长的影响很大; Hoyer 算法使用的两优化方法的特点决定了其效率较低. Liu 等为了解决 Hoyer 算法性能与优化参数相关联的问题, 构造了无用户定义优化参数的稀疏非负矩阵分解 (Sparse NMF, SNMF) 算法^[11] 来实现 NSC. SNMF 算法从数学构造上看很完美, 但实用中, 它并不能达到 Hoyer 算法可达到的效果 (SNMF 算法的解远差于 Hoyer 算法的解, 详细分析见第 2 节).

为增强 NSC 的可应用性, 本文通过对一系列凸超物面函数做交替最小化来实现 NSC, 依据凸超抛物面特性、点到非负数集合的投影规则和点到原点处单位超球的投影规则, 构造一个无用户定义优化参数的稳定高效的 NSC 算法 — SENS. 除比 Hoyer 算法稳定高效外, SENS 的解也会优于 Hoyer 算法的解.

本文内容安排如下: 第 1 节简述 NSC, 说明 Hoyer 算法和 SNMF 算法的缺陷; 第 2 节构造 SENS; 第 3 节解释推断 SENS 相比 Hoyer 算法高效且可得到更优解的原因, 证明 SENS 的稳定性和收敛性; 第 4 节构造一些实验去验证和分析 SENS 的性质; 第 5 节总结全文.

1 NSC 及其算法

对一个 M 维的随机向量 \mathbf{v} 进行 N 次观测, 记这些观测为向量集 $\{\mathbf{v}_j, j = 1, 2, \dots, N\}$, 取 $V = [\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_N]$, 其中 $\mathbf{V}_j = \mathbf{v}_j, j = 1, 2, \dots, N$.

定义 1. NSC 是通过解

$$\min_{W, H \geq 0} f_V(W, H) + \lambda S(H) \quad (1)$$

求得非负的 $M \times L$ 维基矩阵 W 和 $L \times N$ 维系数矩阵 H , 使

$$V \approx WH \quad (2)$$

且使 H 具有一定稀疏性 ($f_V(W, H)$ 度量 V 和 WH 间的差异, $S(H)$ 是为使 H 具有稀疏性而施加的罚项, λ 是一个正常数, 用它可调节编码 H 的稀疏程度, L 表示进行 NSC 后数据的维数).

Hoyer 选择了如下的优化模型^[2]:

$$\min_{W, H \geq 0, \forall i, \|W_{\cdot i}\|_2=1} \|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij} \quad (3)$$

采用交替优化的策略, 对 W 用梯度投影法优化 (这里对 W 的优化是一个很难的问题, 文献 [2] 中称尚无更有效的优化方法), 对 H 用与文献 [12] 中类似的基于辅助函数的乘性更新法优化, 得到了算法 1.

算法 1 (Hoyer's NSC algorithm).

输入. V, L ;

输出. W, H ;

1. 随机初始化 $W(0) \geq 0$ 和 $H(0) \geq 0, \forall i$,
 $W_{\cdot i}(0) \leftarrow W_{\cdot i}(0) / \|W_{\cdot i}(0)\|_2, k \leftarrow 0$;
2. 重复如下操作, 直到满足收敛条件
3. $W' \leftarrow W(t) - \mu[W(t)H(t) - V]H(t)^T$;
4. $W' \leftarrow \max\{W', 0\}$;
5. $\forall i, W_{\cdot i}(t+1) \leftarrow W'_{\cdot i} / \|W'_{\cdot i}\|_2$;
6. $H(t+1) \leftarrow H(t) \otimes [W(t+1)^T V] \odot [W(t+1)^T W(t+1)H(t) + \lambda]$
(\otimes 和 \odot 分别代表 Hardward 乘和除);
7. $t \leftarrow t+1$.

梯度投影法的特点决定了 Hoyer 算法的性能受迭代步长 μ 的影响很大, 取 L 和 λ 均为 100, 图 1 以 ORL 数据库^[13] 的 400 幅图像为实验数据 (为使实验结果具有可比性, 本文中所有实验均基于这组数据进行), 对 Hoyer 算法的这个缺点做了直观说明 (本文中所有实验结果中的计算时间均借助配置为 Pentium 4 2.8 G, 768 M RAM 的 PC 机记录).

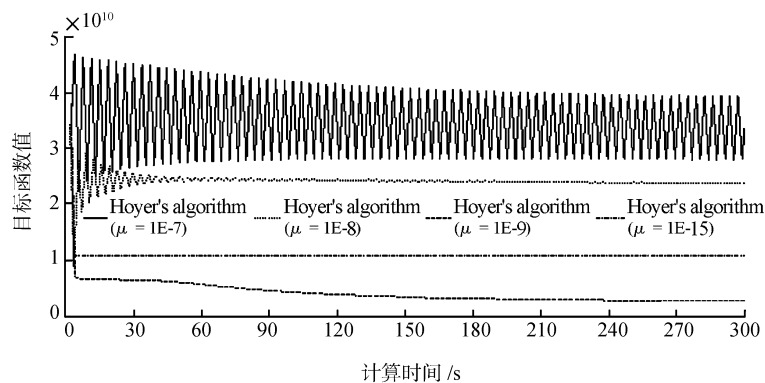


图 1 Hoyer 算法性能与迭代步长 μ 的关系 ($L = 100, \lambda = 100$)

Fig. 1 Relationship between the performance of Hoyer's algorithm and iterative step size μ ($L = 100, \lambda = 100$)

如果 μ 取值不够小, 可能导致算法振荡, 如 μ 取 $1E-7$ 时得到的迭代曲线所示; 减小 μ , 可使振荡幅度变小, 如取 $1E-8$ 时的迭代曲线所示; 当 μ 取得足够小, 通过此算法可使目标函数值单调下降, 如 μ 取 $1E-9$ 时的情形. 应该指出, μ 取多少为足够小, 实验前是无法得知的. 如果对 μ 的取值过小, 虽然可保证目标函数值单调下降, 但会使算法的效率急剧下降, 如 μ 取 $1E-15$ 时的情形, 因此 μ 的取值不能太小.

需要说明的是, 即使 μ 的取值合理, Hoyer 算法的效率也不会很高, 如图 1 中 μ 取 $1E-9$ 时的迭代曲线的下降速率所示. 文献 [12] 中已分析说明基于辅助函数的乘性更新法等价于迭代步长自适应调整的梯度下降法. 一般情况下, 梯度投影法的效率低于梯度下降法, 因此, Hoyer 算法的效率至多与迭代步长选择得当时的梯度下降法相当. 此外, 乘性更新法的特点是, 某优化元素一旦在迭代中出现零或非常接近零的值, 那么它在以后的迭代中将很难再取到较大的非零值, 这使 Hoyer 算法很可能停步在不是很好的局部解上.

Liu 等以

$$\min_{W, H \geq 0, \forall i, \|W_{\cdot i}\|_1=1} \sum_{ij} [V_{ij} \log(\frac{V_{ij}}{[WH]_{ij}}) - V_{ij} + [WH]_{ij}] + \alpha \sum_{ij} H_{ij} \quad (4)$$

为优化模型^[11], 也采用交替优化的策略, 对 W 和 H 均用基于辅助函数的乘性更新法优化, 构造了算法 2 (SNMF 算法) 进行 NSC.

算法 2 (SNMF algorithm).

输入. V, L ;
输出. W, H ;

1. 随机初始化 $W(0) \geq 0$ 和 $H(0) \geq 0, \forall i, W_{\cdot i}(0) \leftarrow W_{\cdot i}(0)/\|W_{\cdot i}(0)\|_1, t \leftarrow 0$;
2. 重复如下操作, 直到满足收敛条件
3. $H(t+1) \leftarrow H(t) \otimes \{W^T(t)\{V \times [W(t)H(t)]\}/(1+\alpha)\}$;
4. $W(t+1) \leftarrow W(t) \otimes \{V[W(t)H(t+1)]H^T(t+1)\} \odot [1_{M \times N}H(t+1)]$ ($1_{M \times N}$ 为所有元素均为 1 的 $M \times N$ 矩阵);
5. $\forall i, W_{\cdot i}(t+1) \leftarrow W_{\cdot i}(t+1)/\|W_{\cdot i}(t+1)\|_1$;
6. $t \leftarrow t+1$.

SNMF 算法无用户定义的优化参数, 可使式 (4) 中的目标函数单调下降, 其迭代性能稳定^[11]. 但是, 其对编码稀疏度调节能力非常有限, 且对数据的描述能力很差, 这可用图 2 和图 3 中记录的实验结果说明. 同样以 ORL 库的图像为实验数据, 取 $L = 100$, 图 2 和图 3 分别记录了不同 α 值下广义 Kullback-Leibler 散度 ($\sum_{ij} V_{ij} \log(V_{ij}/[WH]_{ij} - V_{ij} + [WH]_{ij})$) 值和 H 的稀疏度在迭代过程中的变化. 本文中, 向量的稀疏性用

$$sparseness(\mathbf{x}) = \frac{\sqrt{n} - \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}}{\sqrt{n} - 1}, \quad \forall \mathbf{x} \in \mathbf{R}^n \quad (5)$$

来度量^[14]. 如果 \mathbf{x} 中仅有一个元素非零 (最稀疏的情况), 则 $sparseness(\mathbf{x})$ 为 1; 如果 \mathbf{x} 中所有元素的值均相等 (最不稀疏/最平滑的情况), 则 $sparseness(\mathbf{x})$ 为 0. 因为 NSC 关心 H 的行的稀疏性^[2-3], 进行 NSC 后 H 的稀疏度用

$$sparseness(H) = \frac{\sum_{i=1}^L sparseness(H_i)}{L} \quad (6)$$

来度量.

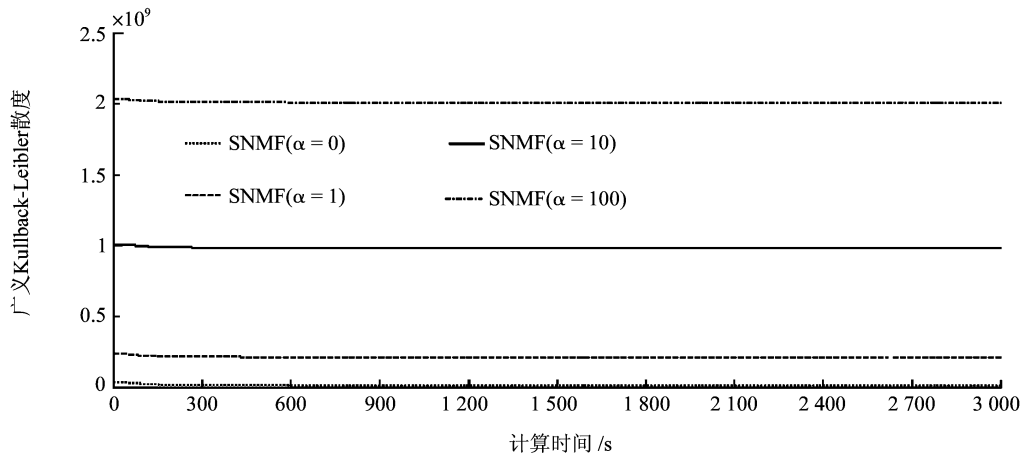


图 2 用 SNMF 算法时, 描述能力与稀疏性调节因子 α 的关系 ($L = 100$)

Fig. 2 Relationship between approximation accuracy and sparse factor α while using SNMF algorithm ($L = 100$)

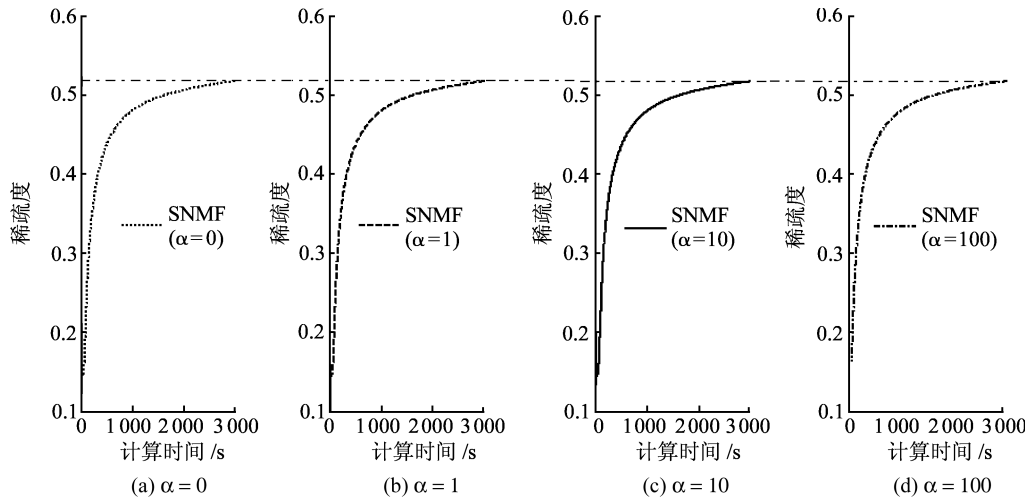


图 3 用 SNMF 算法时, 编码稀疏度与稀疏性调节因子 α 的关系 ($L = 100$)

Fig. 3 Relationship between sparseness degree of codes and sparse factor α while using SNMF algorithm ($L = 100$)

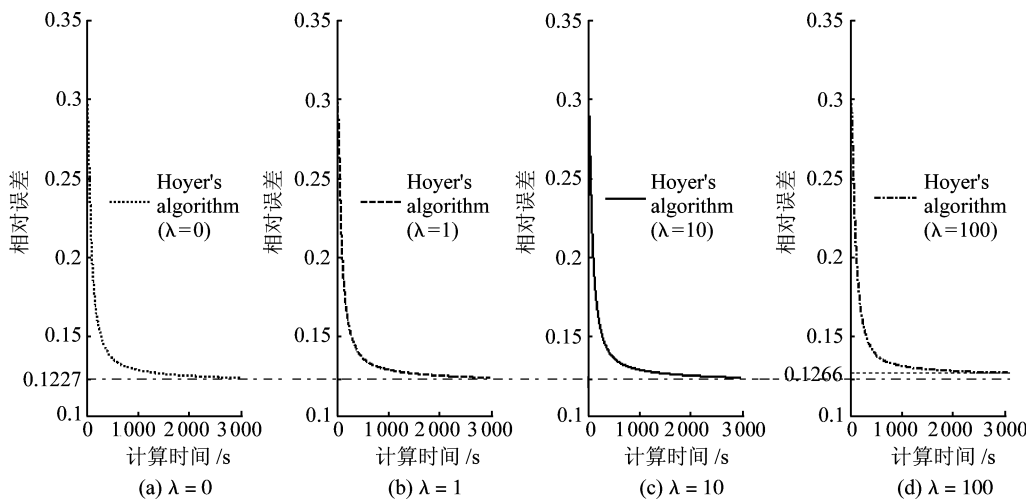


图 4 用 Hoyer 算法时, 描述能力与稀疏性调节因子 λ 的关系 ($L = 100, \mu = 1E-9$)

Fig. 4 Relationship between approximation accuracy and sparse factor λ while using Hoyer's algorithm ($L = 100, \mu = 1E-9$)

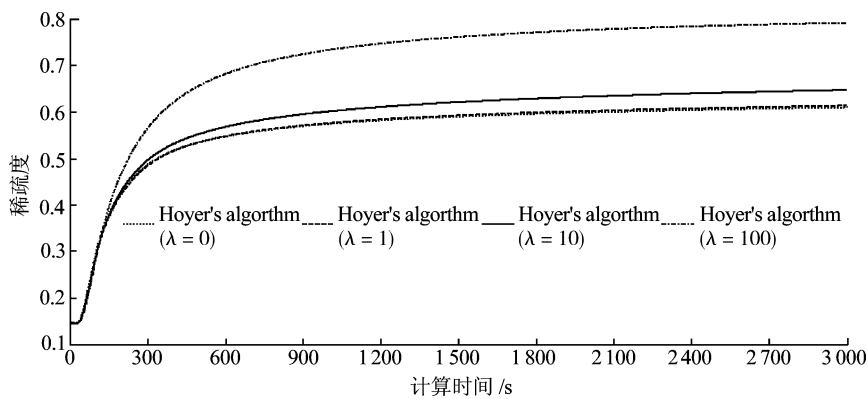


图 5 用 Hoyer 算法时, 编码稀疏度与稀疏性调节因子 λ 的关系 ($L = 100, \mu = 1E-9$)

Fig. 5 Relationship between sparseness degree of codes and sparse factor λ while using Hoyer's algorithm ($L = 100, \mu = 1E-9$)

$\alpha = 0$ 时, SNMF 算法实际上退化为基于最小化 Kullback-Leibler 散度的 NMF 算法^[12]. 观察图 2 知, 当 α 分别取 1、10 和 100 时, 由 SNMF 算法进行的 NSC 对数据的描述能力相比由 SNMF 算法 ($\alpha = 0$) 进行的 NMF 急剧下降. 观察图 3 知, 当 α 分别取 1、10 和 100 时, 由 SNMF 算法进行 NSC 时的编码稀疏度与由 SNMF 算法进行 NMF 时的编码稀疏度相差无几, 也就是说, SNMF 算法调节编码稀疏性的能力很差. SNMF 算法从数学构造上看很完美 (不受优化参数影响, 单调下降), 但实际效果很差, 其原因可能在于式 (4) 中目标函数的空间形状过于复杂, 这使算法极易陷入极差的局部解.

对 Hoyer 算法也进行类似于图 2 和图 3 那样的实验, 其结果记录在图 4 和图 5 中, 这里依图 1 的实验结果取 $\mu = 1\text{E}-9$, 其他实验设置与图 2 和图 3 中的实验相同. $\lambda = 0$ 时, Hoyer 算法退化为基于最小二乘的 NMF 算法, 由图 4 可知, λ 取 1 和 10 时, 由 Hoyer 算法进行的 NSC 对数据的描述能力不比由 Hoyer 算法 ($\lambda = 0$) 进行的 NMF 差, λ 取 100 时, 由 Hoyer 算法进行的 NSC 对数据的描述能力比由 Hoyer 算法 ($\lambda = 0$) 进行的 NMF 稍差. 图 5 说明由 Hoyer 算法进行 NSC 时, 编码稀疏度随 λ 值增大明显提高, 即 Hoyer 算法调节编码稀疏性的能力较强.

综上所述, Hoyer 算法可实现的性能远好于 SNMF 算法, 但其性能受优化参数 (迭代步长) 影响很大, 且效率不高.

2 SENS 的构造

为提高 NSC 的可应用性, 这里构造一个新的 NSC 算法 — SENS.

SENS 仍用 Hoyer 算法采用的优化模型 (见式 (3)), 它将式 (3) 中的目标函数分解为一组凸超抛物面函数, 以对这些凸超抛物面函数做交替最小化为对式 (3) 优化问题的求解策略. 这样的优化策略使 SENS 像 SNMF 算法一样无用户定义的优化参数, 比 Hoyer 算法性能稳定且高效. 此外, SENS 的解通常优于 Hoyer 算法 (具体的算法性质分析见第 3 节).

借助凸超抛物面的特性、点到非负点集的投影规则以及点到原点处单位超球的投影规则, 可构造如下定理, 它们是构造 SENS 的基础.

为证明定理 1, 先证明引理 1 和引理 2.

引理 1. 对于 $\mathbf{x} \in \mathbf{R}^n$, 令 $\gamma \in \mathbf{R}^+$, 则凸超抛物面函数

$$f(\mathbf{x}) = \mathbf{x}^T[\gamma I]\mathbf{x} - 2\mathbf{c}^T\mathbf{x}, \quad I \text{ 为单位阵} \quad (7)$$

的等高超曲面在 \mathbf{x} 所在空间上投影为超球, 它在某

点处的值, 与此点所处超球的半径平方成正线性关系.

证明. 令 $\mathbf{c} = [c_1, c_2, \dots, c_n]^T$, $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$,

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{x}^T[\gamma I]\mathbf{x} - 2\mathbf{c}^T\mathbf{x} = \\ &(\gamma x_1^2 - 2c_1x_1) + (\gamma x_2^2 - 2c_2x_2) + \dots + \\ &(\gamma x_n^2 - 2c_nx_n) = \\ &\left(\sqrt{\gamma}x_1 - \frac{c_1}{\sqrt{\gamma}}\right)^2 + \left(\sqrt{\gamma}x_2 - \frac{c_2}{\sqrt{\gamma}}\right)^2 + \\ &\dots + \left(\sqrt{\gamma}x_n - \frac{c_n}{\sqrt{\gamma}}\right)^2 - \\ &\frac{c_1^2 + c_2^2 + \dots + c_n^2}{\gamma} \end{aligned} \quad (8)$$

显然, $[c_1^2 + c_2^2 + \dots + c_n^2]/\gamma$ 为常数, $\forall r \in \mathbf{R}^+$, 取 $(\sqrt{\gamma}x_1 - c_1/\sqrt{\gamma})^2 + (\sqrt{\gamma}x_2 - c_2/\sqrt{\gamma})^2 + \dots + (\sqrt{\gamma}x_n - c_n/\sqrt{\gamma})^2 = r^2$, 则 $f(\mathbf{x}) = r^2 - [c_1^2 + c_2^2 + \dots + c_n^2]/\gamma$. 所以, 凸超抛物面 $f(\mathbf{x}) = \mathbf{x}^T[\gamma I]\mathbf{x} - 2\mathbf{c}^T\mathbf{x}$ 的等高超曲面在 \mathbf{x} 所在空间上投影为超球, 并且, 它在某点处的值, 与此点所处超球的半径平方成正线性关系. \square

图 6 是 $f(\mathbf{x})$ 的空间形状在可视的三维空间中的示例, 此时, $f(\mathbf{x})$ 的等高超曲面退化为等高曲线 (图 6 空间中的圆线), 这些曲线在 \mathbf{x} 所处空间上的投影为圆 (\mathbf{x} 所在平面上的圆线).

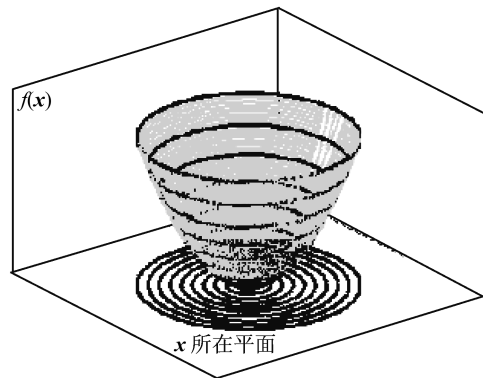


图 6 $f(\mathbf{x})$ 的空间形状在可视的三维空间中的图示
Fig. 6 A drawing example for the characteristics of $f(\mathbf{x})$ in visualized three-dimension space

引理 2 (点到非负点集的投影规则). $\forall \mathbf{a} \in \mathbf{R}^n$, 最小化问题 (找与 \mathbf{a} 最近的非负点)

$$\min_{\mathbf{x} \geq 0} \|\mathbf{x} - \mathbf{a}\|_2^2 \quad (9)$$

的解为

$$\mathbf{x} = \max\{\mathbf{a}, 0\} \quad (10)$$

证明. 令 $\mathbf{a} = [a_1, a_2, \dots, a_n]^T$, $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$,

$$\min_{\mathbf{x} \geq 0} \|\mathbf{x} - \mathbf{a}\|_2^2 \Leftrightarrow \min_{x_1, x_2, \dots, x_n \geq 0} \{(x_1 - a_1)^2 + (x_2 - a_2)^2 + \dots + (x_n - a_n)^2\} \quad (11)$$

式 (11) 中各项的值均达到最小, 则式 (11) 的值达到最小. 按此思路, $\forall i \in \{1, 2, \dots, n\}$, 如果 $a_i \geq 0$, 则最小值点处 $x_i = a_i$, 如果 $a_i \leq 0$, 则最小值点处 $x_i = 0$. \square

定理 1. $\forall j \in \{1, 2, \dots, L\}$, 记

$$F_{ab} = [W^T W]_{ab}(1 - \delta_{aj}\delta_{jb}), \forall a, b \in \{1, 2, \dots, L\} \\ (i = j, \text{ 则 } \delta_{ij} = 1; i \neq j, \text{ 则 } \delta_{ij} = 0) \quad (12)$$

固定 W 和 $\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L\}/\mathbf{H}_j$, 并保证 $\|\mathbf{W}_{\cdot j}\|_1 \geq 0$, 式 (13) 中的目标函数在

$$\mathbf{H}_j = \max \left\{ \frac{[W^T V]_{j\cdot} - \mathbf{F}_j \cdot \mathbf{H} - \lambda}{[W^T W]_{jj}}, 0 \right\} \quad (13)$$

处达到对 \mathbf{H}_j 有非负限制时的最小值, 且式 (13) 所处境点为唯一的最小值点 (向量减标量是指向量的每个元素均减去这个标量).

证明. 因为

$$\begin{aligned} \|V - WH\|_F^2 + 2\lambda \sum_{m=1}^L \sum_{n=1}^N H_{mn} &= \\ \text{tr}\{(V - WH)(V - WH)^T\} + & \\ 2\lambda \sum_{m=1}^L \sum_{n=1}^N H_{mn} &= \\ \text{tr}\{VV^T - 2HV^T W + H^T W^T W H\} + & \\ 2\lambda \sum_{m=1}^L \sum_{n=1}^N H_{mn} &= \\ \text{tr}\{VV^T\} - 2 \sum_{g=1}^N [V^T W]_{g\cdot} \mathbf{H}_{\cdot g} + & \\ \sum_{g=1}^N \mathbf{H}_{\cdot g}^T W^T W \mathbf{H}_{\cdot g} + 2\lambda \sum_{m=1}^L \sum_{n=1}^N H_{mn} & \quad (14) \end{aligned}$$

$\forall f \in \{1, 2, \dots, N\}$, 固定 W 后, 关于 $\mathbf{H}_{\cdot f}$ 的目标函数简化为

$$\begin{aligned} f_H(\mathbf{H}_{\cdot f}) &= \mathbf{H}_{\cdot f}^T W^T W \mathbf{H}_{\cdot f} - 2[V^T W]_{f\cdot} \mathbf{H}_{\cdot f} + \\ 2\lambda \sum_{m=1}^L H_{mf} &= \sum_{k=1}^L \sum_{l=1}^L H_{kf} [W^T W]_{kl} H_{lf} - \\ 2 \sum_{h=1}^L [V^T W]_{fh} H_{hf} + 2\lambda \sum_{m=1}^L H_{mf} & \quad (15) \end{aligned}$$

进一步, 对 $\forall j \in \{1, 2, \dots, L\}$, 固定 W 和 $\{\mathbf{H}_{1f}, \mathbf{H}_{2f}, \dots, \mathbf{H}_{Lf}\}/\mathbf{H}_{jf}$ 后, 关于 \mathbf{H}_{jf} 的目标函数简化为

$$\begin{aligned} f_H(H_{jf}) &= H_{jf}^2 [W^T W]_{jj} + 2H_{jf} \times \\ & \left(\sum_{l=1, l \neq j}^L [W^T W]_{jl} H_{lf} - [V^T W]_{fj} + \lambda \right) \quad (16) \end{aligned}$$

基于式 (12), $\sum_{l=1, l \neq j}^L [W^T W]_{jl} H_{lf} = F_j \cdot H_{\cdot f}$.

由式 (16) 知, $f_H(H_{j1}), f_H(H_{j2}), \dots, f_H(H_{jN})$ 间无关联, 固定 W 和 $\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L\}/\mathbf{H}_j$ 后, 关于 \mathbf{H}_j 的目标函数为

$$\begin{aligned} f_H(\mathbf{H}_{j\cdot}) &= f_H(H_{j1}) + f_H(H_{j2}) + \dots + f_H(H_{jN}) = \\ \mathbf{H}_{j\cdot} \{ [W^T W]_{jj} \mathbf{I} \} \mathbf{H}_{j\cdot}^T - & \\ 2([W^T V]_{j\cdot} - \mathbf{F}_j \cdot \mathbf{H} - \lambda) \mathbf{H}_{j\cdot}^T & \quad (17) \end{aligned}$$

因为 $\|\mathbf{W}_{\cdot j}\|_1 > 0$, $[W^T W]_{jj} > 0$, $f_H(\mathbf{H}_{j\cdot})$ 为凸超抛物面函数, $f_H(\mathbf{H}_{j\cdot})$ 在

$$\frac{[W^T V]_{j\cdot} - \mathbf{F}_j \cdot \mathbf{H} - \lambda}{[W^T W]_{jj}} \quad (18)$$

处达到最小值 (通过 $\frac{\partial f_H(\mathbf{H}_{j\cdot})}{\partial \mathbf{H}_{j\cdot}} = 0$ 计算).

由引理 1 知, $\mathbf{H}_{j\cdot}$ 所处的位置距式 (18) 所在点越近 (所在超球的半径越小), 其对应的目标函数值越小. 因此, 与式 (18) 所处点距离最近的非负点是 $f_H(\mathbf{H}_{j\cdot})$ (固定 W 和 $\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L\}/\mathbf{H}_j$ 时式 (3) 中的目标函数) 在对 $\mathbf{H}_{j\cdot}$ 有非负限制时的最小值点, 由引理 2 知, 此点在式 (13) 处. 因为 $f_H(\mathbf{H}_{j\cdot})$ 是严格凸的 (见式 (17)), 所以它的最小值点是唯一的. \square

推论 1. $\forall j \in \{1, 2, \dots, L\}$, 固定 W 和 $\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L\}/\mathbf{H}_j$, 并保证 $\|\mathbf{W}_{\cdot j}\|_1 > 0$, 在 $\mathbf{H}_{j\cdot} \geq \epsilon$ 的约束下, 式 (3) 中的目标函数在

$$\mathbf{H}_j = \max \left\{ \frac{[W^T V]_{j\cdot} - \mathbf{F}_j \cdot \mathbf{H} - \lambda}{[W^T W]_{jj}}, \epsilon \right\} \quad (19)$$

处达到最小值.

为证明定理 2, 除用到引理 1 和引理 2 外, 还需下面的引理 3 和引理 4.

定义 2. 如果 $Q(\mathbf{h}, \mathbf{g}) \geq F(\mathbf{h})$, 当且仅当 $\mathbf{g} = \mathbf{h}$ 时有 $Q(\mathbf{h}, \mathbf{g}) = F(\mathbf{h})$, 则 $Q(\mathbf{h}, \mathbf{g})$ 称为 $F(\mathbf{h})$ 的辅助函数^[12].

引理 3. 如果 $Q(\mathbf{h}, \mathbf{g})$ 为 $F(\mathbf{h})$ 的辅助函数, 且 \mathbf{g} 固定后, $Q(\mathbf{h}, \mathbf{g})$ 是 \mathbf{h} 的严格凸函数, 则依照

$$\mathbf{h}^{t+1} = \arg \min_{\mathbf{h}} Q(\mathbf{h}, \mathbf{h}^t) \quad (20)$$

调整 \mathbf{h} , 可使 $F(\mathbf{h})$ 严格单调下降 (\mathbf{h}^t 和 \mathbf{h}^{t+1} 分别为 \mathbf{h} 的当前值和更新后的值).

证明. $F(\mathbf{h}^{t+1}) \leq F(\mathbf{h}^{t+1}, \mathbf{h}^t) < G(\mathbf{h}^t, \mathbf{h}^t) = F(\mathbf{h}^t)$. \square

图 7 是对引理 3 的一个直观解释^[12].

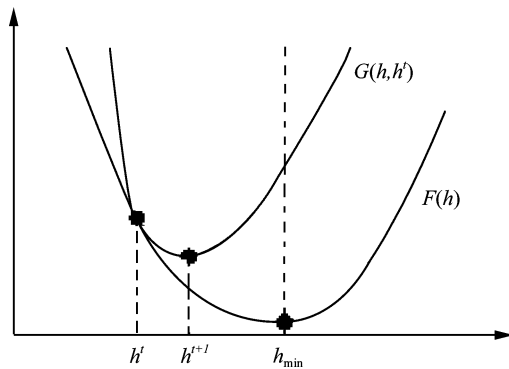


图 7 对引理 3 的图示说明

Fig. 7 A drawing explanation for Lemma 3

引理 4 (点到原点处单位超球的投影规则). $\forall \mathbf{d} \in \mathbf{R}^n$ (其空间位置记为点 a), 最小化问题 (求原点处的单位超球上与点 a 距离最近的点)

$$\min_{\|\mathbf{x}\|_2=1} \|\mathbf{x} - \mathbf{d}\|_2^2 \quad (21)$$

的解为

$$\mathbf{x} = \frac{\mathbf{d}}{\|\mathbf{d}\|_2} \quad (22)$$

证明. 由原点出发指向 a 点的射线与原点处单位超球的交点即是所求之点, 这可通过图 8 从几何上说明, 记交点为 c , 在超球上任取一点 b , 则 a 、 b 和 c 构成一平面, 此平面与超球的交线为一个如图 8 所示的圆, 显然, $ab > ac$. 因为点 b 是任取的, 所以在原点处的单位超球上找不到比点 c 更接近 a 的点. 式 (22) 就是交点 c 的坐标, 因为式 (22) 既满足上述射线方程又满足原点处的单位超球方程, 并且上述的交点是唯一的. \square

定理 2. $\forall i \in \{1, 2, \dots, L\}$, 固定 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\} / \mathbf{W}_{\cdot i} \geq 0$ 和 $\mathbf{H} \geq 0$, 并保证 $\|\mathbf{H}_{\cdot i}\|_1 > 0$,

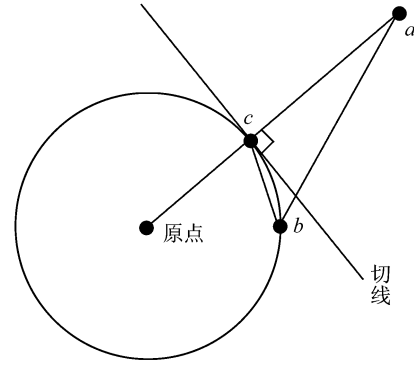


图 8 对点到原点处单位超球投影规则的二维图解
Fig. 8 A two-dimensional drawing explanation for projection formula from a point to the unit super sphere at origin

在 $\mathbf{W}_{\cdot i} \geq 0$ 和 $\|\mathbf{W}_{\cdot i}\| = 1$ 的约束下, 通过

$$\mathbf{W}_{\cdot i} \leftarrow \frac{\mathbf{W}_{\cdot i} \otimes [\mathbf{V}\mathbf{H}^T]_{\cdot i} \odot [\mathbf{W}\mathbf{H}\mathbf{H}^T]_{\cdot i}}{\|\mathbf{W}_{\cdot i} \otimes [\mathbf{V}\mathbf{H}^T]_{\cdot i} \odot [\mathbf{W}\mathbf{H}\mathbf{H}^T]_{\cdot i}\|_2} \quad (23)$$

可使式 (3) 中的目标函数下降.

证明. 类似于式 (14), 有

$$\begin{aligned} \|V - \mathbf{W}\mathbf{H}\|_F^2 &= \\ \text{tr}(\mathbf{V}\mathbf{V}^T - 2\mathbf{W}\mathbf{H}\mathbf{V}^T + \mathbf{W}\mathbf{H}\mathbf{H}^T\mathbf{W}) &= \\ \text{tr}\{\mathbf{V}\mathbf{V}^T\} - 2 \sum_{d=1}^M \mathbf{W}_d [\mathbf{H}\mathbf{V}^T]_{\cdot d} + & \\ \sum_{d=1}^M \mathbf{W}_d \mathbf{H}\mathbf{H}^T \mathbf{W}_d^T & \end{aligned} \quad (24)$$

那么, $\forall c \in \{1, 2, \dots, M\}$, 固定 \mathbf{H} , 关于 \mathbf{W}_c 的目标函数简化为

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{W}_c) &= \mathbf{W}_c \mathbf{H}\mathbf{H}^T \mathbf{W}_c^T - 2\mathbf{W}_c [\mathbf{H}\mathbf{V}^T]_{\cdot c} = \\ & \sum_{k=1}^L \sum_{l=1}^L \mathbf{W}_{ck} [\mathbf{H}\mathbf{H}^T]_{kl} \mathbf{W}_{cl} - \\ & 2 \sum_{h=1}^L \mathbf{W}_{ch} [\mathbf{H}\mathbf{V}^T]_{hc} \end{aligned} \quad (25)$$

进一步, $\forall i \in \{1, 2, \dots, L\}$, 固定 \mathbf{H} 和 $\{\mathbf{W}_{c1}, \mathbf{W}_{c2}, \dots, \mathbf{W}_{cL}\} / \mathbf{W}_{ci}$ 后, 关于 \mathbf{W}_{ci} 的目标函数简化为

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{W}_{ci}) &= [\mathbf{H}\mathbf{H}^T]_{ii} \mathbf{W}_{ci}^2 + 2\mathbf{W}_{ci} \times \\ & \left(\sum_{k=1, k \neq i}^L \mathbf{W}_{ck} [\mathbf{H}\mathbf{H}^T]_{ki} - [\mathbf{V}\mathbf{H}^T]_{ci} \right) \end{aligned} \quad (26)$$

记

$$E_{ab} = [HH^T]_{ab}(1 - \delta_{ai}\delta_{ib}), \quad \forall a, b \in \{1, 2, \dots, L\} \quad (27)$$

有 $\sum_{k=1, k \neq i}^L W_{ck} [HH^T]_{ki} = \mathbf{W}_c \cdot \mathbf{E}_i$. 构造

$$Q_W(W_{ci}, B_{ci}) = f_W(B_{ci}) + (W_{ci} - B_{ci})^2 \times \frac{\{\mathbf{W}_c \cdot \mathbf{E}_i + B_{ci} [HH^T]_{ii}\}}{B_{ci}} + 2(W_{ci} - B_{ci}) \times \{\mathbf{W}_c \cdot \mathbf{E}_i + B_{ci} [HH^T]_{ii} - [VH^T]_{ci}\} \quad (28)$$

显然, $f_W(W_{ci}) = Q_W(W_{ci}, W_{ci})$. 可推知

$$f_W(W_{ci}) = [HH^T]_{ii} W_{ci}^2 + 2W_{ci}(\mathbf{W}_c \cdot \mathbf{E}_i - [VH^T]_{ci}) = [HH^T]_{ii} B_{ci}^2 + 2B_{ci}(\mathbf{W}_c \cdot \mathbf{E}_i - [VH^T]_{ci}) + [HH^T]_{ii} W_{ci}^2 + [HH^T]_{ii} B_{ci}^2 - 2[HH^T]_{ii} W_{ci} B_{ci} - 2[HH^T]_{ii} B_{ci}^2 + 2[HH^T]_{ii} W_{ci} B_{ci} + 2W_{ci}(\mathbf{W}_c \cdot \mathbf{E}_i - [VH^T]_{ci}) - 2B_{ci}(\mathbf{W}_c \cdot \mathbf{E}_i - [VH^T]_{ci}) = f_W(B_{ci}) + [HH^T]_{ii}(W_{ci} - B_{ci})^2 + 2(W_{ci} - B_{ci})(\mathbf{W}_c \cdot \mathbf{E}_i + [HH^T]_{ii} B_{ci} - [VH^T]_{ci}) \quad (29)$$

依题设条件, $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\} / \mathbf{W}_{\cdot i} \geq 0$ 以及 $H \geq 0$, 并不失一般性, 令 $B_{ci} \geq 0$, 则

$$Q_W(W_{ci}, B_{ci}) - f_W(W_{ci}) = (W_{ci} - B_{ci})^2$$

$$\frac{\mathbf{W}_c \cdot \mathbf{E}_i + B_{ci} [HH^T]_{ii}}{B_{ci}} - [HH^T]_{ii} = \frac{(W_{ci} - B_{ci})^2 \mathbf{W}_c \cdot \mathbf{E}_i}{B_{ci}} \geq 0$$

所以, $Q_W(W_{ci}, B_{ci})$ 是 $f_W(W_{ci})$ 的辅助函数.

由式 (26) 知, $f_W(W_{1i}), f_W(W_{2i}), \dots, f_W(W_{Mi})$ 间无关联, 固定 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\} / \mathbf{W}_{\cdot i} \geq 0$ 和 $H \geq 0$ 后, 有关 $\mathbf{W}_{\cdot i}$ 的目标函数化简为

$$f_W(\mathbf{W}_{\cdot i}) = f_W(W_{1i}) + f_W(W_{2i}) + \dots + f_W(W_{Mi}) \quad (30)$$

根据定义 2, $f_W(\mathbf{W}_{\cdot i})$ 的辅助函数为

$$Q_W(W_{\cdot i}, B_{\cdot i}) = Q_W(W_{1i}, B_{1i}) + Q_W(W_{2i}, B_{2i}) + \dots + Q_W(W_{Mi}, B_{Mi}) \quad (31)$$

因为 $\|\mathbf{H}_i\|_1 > 0, \{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\} / \mathbf{W}_{\cdot i} \geq 0$ 以及 $H \geq 0$, 并不失一般性令 $\mathbf{B}_i > 0$, 所以 $\forall c$,

$\{\mathbf{W}_c \cdot \mathbf{E}_i + B_{ci} [HH^T]_{ii}\} / B_{ci} > 0, Q_W(\mathbf{W}_{\cdot i}, \mathbf{B}_i)$ 为严格凸超抛物面函数, 其在满足

$$\frac{\partial Q_W(\mathbf{W}_{\cdot i}, \mathbf{B}_i)}{\partial \mathbf{W}_{\cdot i}} = \begin{bmatrix} \frac{\partial Q_W(W_{1i}, B_{1i})}{\partial W_{1i}} \\ \frac{\partial Q_W(W_{2i}, B_{2i})}{\partial W_{2i}} \\ \vdots \\ \frac{\partial Q_W(W_{Mi}, B_{Mi})}{\partial W_{Mi}} \end{bmatrix} =$$

$$\begin{bmatrix} 2(W_{1i} - B_{1i}) \frac{\{\mathbf{W}_{1\cdot} \cdot \mathbf{E}_i + B_{1i} [HH^T]_{ii}\}}{B_{1i}} + 2(\mathbf{W}_{1\cdot} \cdot \mathbf{E}_i + [HH^T]_{ii} B_{1i} - [VH^T]_{1i}) \\ 2(W_{2i} - B_{2i}) \frac{\{\mathbf{W}_{2\cdot} \cdot \mathbf{E}_i + B_{2i} [HH^T]_{ii}\}}{B_{2i}} + 2(\mathbf{W}_{2\cdot} \cdot \mathbf{E}_i + [HH^T]_{ii} B_{2i} - [VH^T]_{2i}) \\ \vdots \\ 2(W_{Mi} - B_{Mi}) \frac{\{\mathbf{W}_{M\cdot} \cdot \mathbf{E}_i + B_{Mi} [HH^T]_{ii}\}}{B_{Mi}} + 2(\mathbf{W}_{M\cdot} \cdot \mathbf{E}_i + [HH^T]_{ii} B_{Mi} - [VH^T]_{Mi}) \end{bmatrix} = 0 \quad (32)$$

的点处达到最小值, 易知此点为

$$\mathbf{B}_i \otimes [VH^T]_{\cdot i} \odot \{\mathbf{W} \mathbf{E}_i + [HH^T]_{ii} \mathbf{B}_i\} \quad (33)$$

根据引理 1 和引理 4, $Q_W(\mathbf{W}_{\cdot i}, \mathbf{B}_i)$ 在 $\|\mathbf{W}_{\cdot i}\| = 1$ 的约束下, 于

$$\frac{\mathbf{B}_i \otimes [VH^T]_{\cdot i} \odot \{\mathbf{W} \mathbf{E}_i + [HH^T]_{ii} \mathbf{B}_i\}}{\|\mathbf{B}_i \otimes [VH^T]_{\cdot i} \odot \{\mathbf{W} \mathbf{E}_i + [HH^T]_{ii} \mathbf{B}_i\}\|_2} \quad (34)$$

处达到最小值. 取 \mathbf{B}_i 为当前的 $\mathbf{W}_{\cdot i}$, 式 (34) 就是式 (23) 的右半部分 (注意 $\mathbf{W} \mathbf{E}_i + [HH^T]_{ii} \mathbf{W}_{\cdot i} = [W H H^T]_{\cdot i}$). 根据引理 3 知, 按式 (23) 更新 $\mathbf{W}_{\cdot i}$ 使式 (30) (固定 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\} / \mathbf{W}_{\cdot i} \geq 0$ 和 $H \geq 0$ 时, 式 (3) 中的目标函数) 下降. 依题设条件, 式 (23) 中右端均为非负数, 所以更新后的 $\mathbf{W}_{\cdot i}$ 中也无负数. \square

根据推论 1 和定理 2, 给 W 取定非负的初值, 给 H 取定大于 ϵ 的初值, 对 $\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}, \mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L$ 进行交替优化可在满足式 (3) 中的约束条件的情况下使其目标函数严格单调下降. 定理 2 给出的迭代规则属于乘性更新法, 乘性更新法的缺点在第 1 节中已作说明. 当满足一定的条件 (见定理 3 的题设条件) 时, 可依据如下的定理 3 更新 $\forall i, \mathbf{W}_{\cdot i}$ 以得到更优的解 (原因在第 3 节中说明).

定理 3. $\forall i \in \{1, 2, \dots, L\}$, 固定 H 和 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\}/\mathbf{W}_{\cdot i}$, 并保证 $\|\mathbf{H}_i\|_1 > 0$, 如果 $\|\max\{([VH^T]_{\cdot i} - W\mathbf{E}_i)/[HH^T]_{ii}, 0\}\|_2 \geq 1$, 在 $\mathbf{W}_{\cdot i} \geq 0$ 和 $\|\mathbf{W}_{\cdot i}\| = 1$ 的约束下, 通过

$$\mathbf{W}_{\cdot i} \leftarrow \frac{\max\left\{\frac{[VH^T]_{\cdot i} - W\mathbf{E}_i}{[HH^T]_{ii}}, 0\right\}}{\left\|\max\left\{\frac{[VH^T]_{\cdot i} - W\mathbf{E}_i}{[HH^T]_{ii}}, 0\right\}\right\|_2} \quad (35)$$

可使式 (3) 中的目标函数下降 (\mathbf{E} 与式 (27) 中定义同).

证明. 依据式 (26) 和 (30), 固定 H 和 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\}/\mathbf{W}_{\cdot i}$, 并保证 $\|\mathbf{H}_i\|_1 > 0$

$$f_W(\mathbf{W}_{\cdot i}) = \mathbf{W}_{\cdot i}^T \{[HH^T]_{ii} I\} \mathbf{W}_{\cdot i} - 2\mathbf{W}_{\cdot i}^T \{[VH^T]_{\cdot i} - W\mathbf{H}_i\} \quad (36)$$

为凸超抛物面函数, 它在

$$([VH^T]_{\cdot i} - W\mathbf{E}_i)/[HH^T]_{ii} \quad (37)$$

处达到最小值 (通过 $\frac{\partial f_W(\mathbf{W}_{\cdot i})}{\partial \mathbf{W}_{\cdot i}} = 0$ 计算).

如果 $\|\max\{([VH^T]_{\cdot i} - W\mathbf{E}_i)/[HH^T]_{ii}, 0\}\|_2 \geq 1$, 记 $([VH^T]_{\cdot i} - W\mathbf{E}_i)/[HH^T]_{ii}$ 中大于零元素的行下标集为 $O = \{k \mid ([VH^T]_{ki} - W\mathbf{E}_{ki})/[HH^T]_{ii} > 0\}$, \mathbf{W}_{O_i} 为由 $\mathbf{W}_{\cdot i}$ 按下标集 O 提取出的向量, \mathbf{W}_{P_i} 中除 \mathbf{W}_{O_i} 外的元素组成的向量记为 \mathbf{W}_{P_i} , 显然, $P = \{k \mid ([VH^T]_{ki} - W\mathbf{E}_{ki})/[HH^T]_{ii} \leq 0\}$. \mathbf{W}_{O_i} 的目标函数为

$$f_W(\mathbf{W}_{O_i}) = \mathbf{W}_{O_i}^T \{[HH^T]_{ii} I\} \mathbf{W}_{O_i} - 2\mathbf{W}_{O_i}^T \{[VH^T]_{O_i} - W_O \mathbf{E}_i\} \quad (38)$$

其中, W_O 为按 O 中元素对应行从 W 提取出的矩阵.

\mathbf{W}_{P_i} 的目标函数为

$$f_W(\mathbf{W}_{P_i}) = \mathbf{W}_{P_i}^T \{[HH^T]_{ii} I\} \mathbf{W}_{P_i} - 2\mathbf{W}_{P_i}^T \{[VH^T]_{P_i} - W_P \mathbf{E}_i\} \quad (39)$$

其中, W_P 为按 P 中元素对应行从 W 提取出的矩阵.

由式 (26) 和 (30) 知, $f_W(\mathbf{W}_{O_i})$ 和 $f_W(\mathbf{W}_{P_i})$ 间无关联, 其分别下降可使 $f_W(\mathbf{W}_{\cdot i})$ 下降.

显然, $f_W(\mathbf{W}_{O_i})$ 和 $f_W(\mathbf{W}_{P_i})$ 均为凸超抛物面. 依据引理 1 和引理 2, $f_W(\mathbf{W}_{P_i})$ 在 $\mathbf{W}_{P_i} \geq 0$ 的约束下在零点处达到最小值. $f_W(\mathbf{W}_{O_i})$ 在

$$\frac{[VH^T]_{O_i} - W_O \mathbf{E}_i}{[HH^T]_{ii}} \quad (40)$$

处达到最小值 (通过 $\frac{\partial f_W(\mathbf{W}_{O_i})}{\partial \mathbf{W}_{O_i}} = 0$ 计算). 依据引理 1 和引理 4, $f_W(\mathbf{W}_{O_i})$ 在 $\mathbf{W}_{O_i} > 0$ 和 $\|\mathbf{W}_{O_i}\| = 1$ 的约束下, 于

$$\frac{\frac{[VH^T]_{O_i} - W_O \mathbf{E}_i}{[HH^T]_{ii}}}{\left\|\frac{[VH^T]_{O_i} - W_O \mathbf{E}_i}{[HH^T]_{ii}}\right\|_2} \quad (41)$$

处达到最小值. 依题设条件, 有 $([VH^T]_{O_i} - W_O \mathbf{E}_i)/[HH^T]_{ii} > 1$, $([VH^T]_{O_i} - W_O \mathbf{E}_i)/[HH^T]_{ii}$ 到 $\{\mathbf{W}_{O_i} \mid \|\mathbf{W}_{O_i}\|_2 = 1\}$ 的距离一定小于到 $\{\mathbf{W}_{O_i} \mid \|\mathbf{W}_{O_i}\|_2 < 1\}$ 的距离, 根据引理 1 知, $f_W(\mathbf{W}_{O_i})$ 在 $\mathbf{W}_{O_i} > 0$ 和 $\|\mathbf{W}_{O_i}\| = 1$ 的约束下的最小值一定小于在 $\mathbf{W}_{O_i} > 0$ 和 $\|\mathbf{W}_{O_i}\| < 1$ 的约束下的最小值. 而在 $\mathbf{W}_{O_i} \geq 0$ 和 $\|\mathbf{W}_{O_i}\| = 1$ 的约束下, \mathbf{W}_{O_i} 的当前值一定满足 $\|\mathbf{W}_{O_i}\| \leq 1$, 所以式 (41) 较 \mathbf{W}_{O_i} 的当前值使 $f_W(\mathbf{W}_{O_i})$ 下降 (至少不升).

在 $\mathbf{W}_{P_i} \geq 0$ 的约束下 $f_W(\mathbf{W}_{P_i})$ 的最小值可与在 $\mathbf{W}_{O_i} > 0$ 和 $\|\mathbf{W}_{O_i}\| = 1$ 约束下 $f_W(\mathbf{W}_{O_i})$ 的最小值合并写为

$$\frac{\max\left\{\frac{([VH^T]_{\cdot i} - W\mathbf{E}_i)}{[HH^T]_{ii}}, 0\right\}}{\left\|\max\left\{\frac{([VH^T]_{\cdot i} - W\mathbf{E}_i)}{[HH^T]_{ii}}, 0\right\}\right\|_2} \quad (42)$$

它满足 $\mathbf{W}_{\cdot i} \geq 0$ 且 $\|\mathbf{W}_{\cdot i}\| = 1$ 的约束, 因为这时 $\mathbf{W}_{P_i} = 0$. 式 (42) 是式 (35) 的右半部分, 由上述分析知, 用式 (35) 更新 $\mathbf{W}_{\cdot i}$ 使式 (36) (固定 H 和 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\}/\mathbf{W}_{\cdot i}$ 时, 式 (3) 中的目标函数) 下降. \square

给 W 取定非负的初值, 给 H 取定大于 ϵ 的初值, 交替优化^[15] 中, 根据推论 1 对 $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L$ 进行更新, 当 $\|\max\{([VH^T]_{\cdot i} - W\mathbf{E}_i)/[HH^T]_{ii}, 0\}\|_2 \geq 1$ 时, 根据定理 3 对 $\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}$ 进行更新, 否则根据定理 2 对 $\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}$ 进行更新, 可构造算法 3 (见下页上方).

值得指出, 用推论 1 而不用定理 1 优化 $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L$, 不仅避免了 $\forall ij, [H(k)H^T(k)]_{ij} = 0$, 从而避免了在 SENS 算法中第 6 行和第 8 行出现数学上无定义情况 (除以“0”) 的可能性, 而且几乎不影响算法的下降速度 (因 ϵ 设定为一个足够小的数).

3 SENS 的性质

本节讨论 SENS 的解、效率、稳定性以及收敛性.

算法 3 (SENSC).

输入. V, L ;

输出. W, H ;

1. 随机初始化 $W(0) \geq 0$ 和 $H(0) \geq \varepsilon, \forall i, \mathbf{W}_{\cdot i}(0) \leftarrow \mathbf{W}_{\cdot i}(0)/\|\mathbf{W}_{\cdot i}(0)\|_2, k \leftarrow 0$;

2. 重复如下操作, 直到

$$(\|V - W(k)H(k)\|_F - \|V - W(k+1)H(k+1)\|_F)/\|V\|_F < 10^{-5} \text{ 且} \\ |\sum_{i=1}^L \text{sparseness}\{\mathbf{H}_i(k+1)\}/L - \sum_{i=1}^L \text{sparseness}\{\mathbf{H}_i(k)\}/L| < 10^{-5};$$

3. i 从 1 到 L 执行

4. $E_{ab}(k) = [H(k)H^T(k)]_{ab}(1 - \delta_{ai}\delta_{ib}), \forall a, b \in \{1, 2, \dots, L\}$;

5. 如果 $\|\max\{[VH^T(k)]_{\cdot i} - [\mathbf{W}_{\cdot 1}(k+1), \dots, \mathbf{W}_{\cdot i-1}(k+1), \mathbf{W}_{\cdot i}(k), \dots, \mathbf{W}_{\cdot L}(k)]\mathbf{E}_{\cdot i}\}/[H(k)H^T(k)]_{ii}, 0\}\|_2 \geq 1$, 则

$$6. \quad \mathbf{W}_{\cdot i}(k+1) \leftarrow \frac{\max\left\{\frac{[VH^T(k)]_{\cdot i} - [\mathbf{W}_{\cdot 1}(k+1), \dots, \mathbf{W}_{\cdot i-1}(k+1), \mathbf{W}_{\cdot i}(k), \dots, \mathbf{W}_{\cdot L}(k)]\mathbf{E}_{\cdot i}}{[H(k)H^T(k)]_{ii}}, 0\right\}}{\left\|\max\left\{\frac{[VH^T(k)]_{\cdot i} - [\mathbf{W}_{\cdot 1}(k+1), \dots, \mathbf{W}_{\cdot i-1}(k+1), \mathbf{W}_{\cdot i}(k), \dots, \mathbf{W}_{\cdot L}(k)]\mathbf{E}_{\cdot i}}{[H(k)H^T(k)]_{ii}}, 0\right\}\right\|_2};$$

7. 否则

$$8. \quad \mathbf{W}_{\cdot i}(k+1) \leftarrow \frac{\mathbf{W}_{\cdot i}(k) \otimes [VH^T(k)]_{\cdot i} \odot \{[\mathbf{W}_{\cdot 1}(k+1), \dots, \mathbf{W}_{\cdot i-1}(k+1), \mathbf{W}_{\cdot i}(k), \dots, \mathbf{W}_{\cdot L}(k)]H(k)H^T(k)\}_{\cdot i}}{\|\mathbf{W}_{\cdot i}(k) \otimes [VH^T(k)]_{\cdot i} \odot \{[\mathbf{W}_{\cdot 1}(k+1), \dots, \mathbf{W}_{\cdot i-1}(k+1), \mathbf{W}_{\cdot i}(k), \dots, \mathbf{W}_{\cdot L}(k)]H(k)H^T(k)\}_{\cdot i}\|_2};$$

9. j 从 1 到 L 执行

10. $F_{ab}(k) = [W^T(k+1)W(k+1)]_{ab}(1 - \delta_{aj}\delta_{jb}), \forall a, b \in \{1, 2, \dots, L\}$;

$$11. \quad \mathbf{H}_j(k+1) \leftarrow \max\left\{\frac{[W^T(k+1)V]_j - \mathbf{F}_j(k)[\mathbf{H}_1^T(k+1), \dots, \mathbf{H}_{j-1}^T(k+1), \mathbf{H}_j^T(k), \dots, \mathbf{H}_L^T(k)]^T - \lambda}{[W^T(k+1)W(k+1)]_{jj}}, \varepsilon\right\};$$

12. $k \leftarrow k + 1$.

3.1 SENSIC 的解

Hoyer 算法中, 对 H 的优化使用乘性更新法 (参见第 1 节), 一旦在迭代某步 H 中某元素出现零或非常接近零的值, 那么在以后的迭代中此元素将很难再取到很大的非负值, 这是因为在使用乘性更新法时, 迭代更新值与当前值以乘的方式形成了很强的关联性 (见 Hoyer 算法第 6 行). 这使得 Hoyer 算法在实施过程中 H 的可行域有不断缩小的趋势, 可行域变小往往意味着得到好的解的可能性在减少, 因此, Hoyer 算法很可能停止在不是很好的局部解上.

SENSIC 对 H 使用类似定点算法的优化方式, 即使当前步迭代结果中 H 的某个元素为 ε , 下一步迭代结果中此元素可取到任何非负值.

虽然基于定理 2 的优化方式属于乘性更新法, 但基于定理 3 的优化方式不是乘性更新法, SENSIC 依算法 3 中第 5 行的条件是否满足, 对 W 的优化在上述两种方式间切换, 这保证了即使当前迭代结果中 W 的某元素取到零或非常接近零的值, 经过若干步迭代, 此元素可以取到任何非负值. 而且需要指出的是, 算法 3 的第 5 行的条件并不苛刻, 这个条件可以解释为 $f_W(\mathbf{W}_{\cdot i})$ 在非负约束下的最小值点的模

大于 1, 对绝大多数的数据 (特别是图像数据) 实施 SENSIC 时, 它可被频繁满足.

显然, 在执行 SENSIC 的过程中, W 和 H 的可行域均无缩小的趋势, 因此推断 SENSIC 比 Hoyer 算法易于得到更好的解.

3.2 SENSIC 的效率

根据定理 1 和推论 1, 执行 SENSIC 时, 针对 H 的每次操作 (对 $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L$ 中任一个的更新) 都解析地使当前操作对应的目标函数 (根据当前操作针对的变量, 由整体目标函数化简而得, 如定理 1 证明中化简出的式 (17)) 几乎达到最小值 (只差 ε), 这是典型的定点更新操作. 通过分析文献 [2] 中的推导可知, Hoyer 算法第 6 行, 也可看成是对 $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L$ ($L \leq N$) 逐一进行更新, 它根据基于辅助函数的下降法构造. 由文献 [2] 及引理 3 知, Hoyer 算法能确保对 $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L$ 中任一个的更新都使此操作对应的目标函数下降, 但对下降程度无法控制. 由图 7 的示例可见, 基于辅助函数下降求得的解可能离最小值点很远, 这制约了基于辅助函数的优化法的效率 (图 7 中, 令迭代开始时 \mathbf{h} 为 \mathbf{h}^t , 此时的辅助函数为 $G(\mathbf{h}, \mathbf{h}^t)$, 它的最小值点是 \mathbf{h}^{t+1} , 那么 \mathbf{h}^{t+1} 是此次操作的结果, $F(\mathbf{h}^{t+1})$

$< F(\mathbf{h}^t)$, 但 \mathbf{h}^{t+1} 离 $F(\mathbf{h})$ 的最小值点 \mathbf{h}_{\min} 还有很远的距离). 由以上分析推断, SENS 对 H 的优化效率高于 Hoyer 算法. Hoyer 算法采用梯度投影法更新 W , 为使迭代过程无震荡, 需要设定很小的迭代步长 (第 2 节中已对此用实验说明), 这通常导致较低的效率. SENS 对 W 有两个优化模式, 一个是基于定理 2, 它属于基于辅助函数的乘性更新投影法, 它的投影操作比 Hoyer 算法中的投影操作简单 (比较算法 1 的第 3~5 行和算法 3 的第 8 行可知), 文献 [12] 中已分析, 基于辅助函数的乘性更新法等价于步长自适应的梯度下降法, 步长的自适应避免了固定过小的步长而降低效率, 所以推断用此法对 W 优化的效率会比用梯度投影法高. 另一优化模式基于定理 3, 它把对 \mathbf{W}_i 的优化分为两部分, 一部分针对 \mathbf{W}_{P_i} , 另一部分针对 \mathbf{W}_{O_i} , 每部分都在一定的约束下解析地使其对应目标函数达到最小 (见定理 3 的证明). 这样的优化策略具有定点更新的特点, 其通常比梯度投影法具有强得多的下降能力, 因此, 推断其比梯度投影法效率高 (实际上, 实验中观察其效率往往远高于梯度投影法). 由以上分析推断, SENS 对 W 的优化效率也高于 Hoyer 算法.

综上所述, 推断 SENS 的效率高于 Hoyer 算法.

3.3 SENS 的稳定性

为证明 SENS 稳定性定理 (定理 4), 需如下的引理 5 和引理 6.

引理 5. $\forall A \in \mathbf{R}^{M \times L}, \|A\|_{\text{Abs-sum}} = \sum_{i=1}^M \sum_{j=1}^L |A_{ij}|$ 是 A 的范数.

证明. $\forall A \in \mathbf{R}^{M \times L}, \forall C \in \mathbf{R}^{L \times N},$

$$\begin{aligned} \|AC\|_{\text{Abs-sum}} &= \\ & \left\| \left[\sum_{i=1}^L \mathbf{A}_i C_{i1}, \sum_{i=1}^L \mathbf{A}_i C_{i2}, \dots, \right. \right. \\ & \left. \left. \sum_{i=1}^L \mathbf{A}_i C_{iN} \right] \right\|_{\text{Abs-sum}} = \\ & \sum_{j=1}^N \left\| \sum_{i=1}^L \mathbf{A}_i C_{ij} \right\|_1 \leq \sum_{j=1}^N \sum_{i=1}^L \|\mathbf{A}_i\|_1 |C_{ij}| = \\ & \sum_{i=1}^L \|\mathbf{A}_i\|_1 \|C_i\|_1 \leq \sum_{i=1}^L \|\mathbf{A}_i\|_1 \sum_{i=1}^L \|C_i\|_1 = \\ & \text{(根据 Holder 不等式可得)} \\ & \|A\|_{\text{Abs-sum}} \|C\|_{\text{Abs-sum}} \text{ (相合性)} \quad (43) \end{aligned}$$

$\forall A, B \in \mathbf{R}^{M \times L}, \forall \alpha \in \mathbf{R}, \|A\|_{\text{Abs-sum}} \geq 0$ (非负性), $\|\alpha A\|_{\text{Abs-sum}} = |\alpha| \|A\|_{\text{Abs-sum}}$ (齐次性) 以及 $\|A + B\|_{\text{Abs-sum}} \leq \|A\|_{\text{Abs-sum}} +$

$\|B\|_{\text{Abs-sum}}$ (三角不等式) 是显然成立的. \square

引理 6. 依 SENS 的迭代规则, 序列 $\{(W(k), H(k))\}$ 是有界的.

证明. 假设 $\{(W(k), H(k))\}$ 无界, 因为 $\forall i, \|\mathbf{W}_i(t)\|_2 = 1$, 所以 $\{H(k)\}$ 无界, 那么, $\{\|H(k)\|_{\text{Abs-sum}}\}$ 无界. 因为 $\{\|V - W(k) \times H(k)\|_F^2\}$ 有下界, 所以 $\{\|V - W(k)H(k)\|_F^2 + 2\lambda \times \|H(k)\|_{\text{Abs-sum}}\}$ 无界.

依 SENS 的迭代规则, $\{\|V - W(k)H(k)\|_F^2 + 2\lambda \|H(k)\|_{\text{Abs-sum}}\}$ 单调下降且有下界, 所以它有界, 这与假设 $\{(W(k), H(k))\}$ 无界导出的结论相悖, 所以此假设不成立. \square

定理 4 (SENS 的稳定性定理). 依据 SENS 的迭代原则, $\forall b \in \mathbf{Z}^+, \|(W(k+b), H(k+b)) - (W(k), H(k))\| \rightarrow 0$ (\mathbf{Z}^+ 为正整数集).

证明. 假设 $\|(W(k+1), H(k+1)) - (W(k), H(k))\| \rightarrow 0$ 不成立, 那么存在 $\{(W(k_l), H(k_l))\}$ 的子列 $\{(W(k_{l_h}), H(k_{l_h}))\}$, 使 $\|(W(k_{l_h}+1), H(k_{l_h}+1)) - (W(k_{l_h}), H(k_{l_h}))\| > \delta > 0$ ($l = 1, 2, \dots$).

根据引理 6, $\{(W(k), H(k))\}$ 有界, 所以 $\{(W(k_l), H(k_l))\}$ 有界, 那么存在 $\{(W(k_{l_h}), H(k_{l_h}))\}$ 的收敛子列 $\{(W(k_{l_h_q}), H(k_{l_h_q}))\}$. 显然, $\|(W(k_{l_h_q}+1), H(k_{l_h_q}+1)) - (W(k_{l_h_q}), H(k_{l_h_q}))\| > \delta > 0$ ($h = 1, 2, \dots$).

记 $\{(W(k_{l_h_q}), H(k_{l_h_q}))\} \rightarrow (\bar{W}, \bar{H})$, 那么 $\{\|V - W(k)H(k)\|_F^2 + 2\lambda \sum_{ij} H_{ij}(k)\} \rightarrow \|V - \bar{W}\bar{H}\|_F^2 + 2\lambda \sum_{ij} \bar{H}_{ij}$, 因为: 1) 当 $\forall \bar{W} \in \mathbf{R}^{M \times L}$ 和 $\forall \bar{H} \in \mathbf{R}^{L \times N}$ 时, $\|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij}$ 在 (\bar{W}, \bar{H}) 连续; 2) $\{\|V - W(k)H(k)\|_F^2 + 2\lambda \sum_{ij} H_{ij}(k)\}$ 是收敛的, 它与 $\{\|V - W(k_{l_h_q})H(k_{l_h_q})\|_F^2 + 2\lambda \sum_{ij} H_{ij}(k_{l_h_q})\}$ 收敛到同一点.

根据 SENS 的迭代规则, $\{(\mathbf{W}_{\cdot 1}(k_{l_h_q}+1), \mathbf{W}_{\cdot 2}(k_{l_h_q}), \dots, \mathbf{W}_{\cdot L}(k_{l_h_q}), \mathbf{H}(k_{l_h_q}))\}$ 可以认为是 $\{(W(k), H(k))\}$ 的子列, 所以 $\{\|V - [\mathbf{W}_{\cdot 1}(k_{l_h_q}+1), \mathbf{W}_{\cdot 2}(k_{l_h_q}), \dots, \mathbf{W}_{\cdot L}(k_{l_h_q})]\|_F^2 - 2\lambda \sum_{ij} H(k_{l_h_q})\} \rightarrow \|V - \bar{W}\bar{H}\|_F^2 + 2\lambda \sum_{ij} \bar{H}_{ij}$.

$\{\mathbf{W}_{\cdot 1}(k_{l_h_q}+1)\}$ 有界, 任取 $\{\mathbf{W}_{\cdot 1}(k_{l_h_q}+1)\}$ 中收敛的子列 $\{\mathbf{W}_{\cdot 1}(k_{l_h_q_q}+1)\}$. 因为根据式 (36), 固定 $\mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}$ 和 H 后, $\|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij}$ 关于 $\mathbf{W}_{\cdot 1}$ 严格格凸, 又因 $\{(\mathbf{W}_{\cdot 2}(k_{l_h_q_q}), \dots, \mathbf{W}_{\cdot L}(k_{l_h_q_q}), H(k_{l_h_q_q}))\} \rightarrow (\bar{\mathbf{W}}_{\cdot 2}, \dots, \bar{\mathbf{W}}_{\cdot L}, \bar{H})$, 所以 $\{(\mathbf{W}_{\cdot 1}(k_{l_h_q_q}+1), \mathbf{W}_{\cdot 2}(k_{l_h_q_q}), \dots, \mathbf{W}_{\cdot L}(k_{l_h_q_q}), H(k_{l_h_q_q}))\} \rightarrow (\bar{\mathbf{W}}_{\cdot 1}, \bar{\mathbf{W}}_{\cdot 2}, \dots, \bar{\mathbf{W}}_{\cdot L}, \bar{H})$, 那么 $\{\mathbf{W}_{\cdot 1}(k_{l_h_q_q}+1)\} \rightarrow \bar{\mathbf{W}}_{\cdot 1}$. 因为子列 $\{\mathbf{W}_{\cdot 1}(k_{l_h_q_q}+1)\}$ 为任取的, 所以 $\{\mathbf{W}_{\cdot 1}(k_{l_h_q}+1)\} \rightarrow \bar{\mathbf{W}}_{\cdot 1}$.

那么, $\{(\mathbf{W}_{\cdot 1}(k_{l_h_q}+1), \mathbf{W}_{\cdot 2}(k_{l_h_q}), \dots, \mathbf{W}_{\cdot L}(k_{l_h_q}), H(k_{l_h_q}))\} \rightarrow (\bar{\mathbf{W}}_{\cdot 1}, \bar{\mathbf{W}}_{\cdot 2}, \dots, \bar{\mathbf{W}}_{\cdot L}, \bar{H})$. 以此类推, $\{(\mathbf{W}_{\cdot 1}(k_{l_h}+1), \mathbf{W}_{\cdot 2}(k_{l_h}+1), \dots, \mathbf{W}_{\cdot L}(k_{l_h}),$

$H(k_{l_h}))\} \rightarrow (\bar{W}_{.1}, \bar{W}_{.2}, \dots, \bar{W}_{.L}, \bar{H})$. 由此可见, $\{(W(k_{l_h} + 1), H(k_{l_h}))\} \rightarrow (\bar{W}, \bar{H})$.

同理, $\{(W(k_{l_h} + 1), H(k_{l_h} + 1))\} \rightarrow (\bar{W}, \bar{H})$. 所以, $\|(W(k_{l_h} + 1), H(k_{l_h} + 1)) - (W(k_{l_h}), H(k_{l_h}))\| \rightarrow 0$, 这与假设 $\|(W(k + 1), H(k + 1)) - (W(k), H(k))\| \rightarrow 0$ 不成立推出的结论相悖. 所以, $\|(W(k + 1), H(k + 1)) - (W(k), H(k))\| \rightarrow 0$.

$\|(W(k + 2), H(k + 2)) - (W(k), H(k))\| \leq \|(W(k + 2), H(k + 2)) - (W(k + 1), H(k + 1))\| + \|(W(k + 1), H(k + 1)) - (W(k), H(k))\| \rightarrow 0$. 根据数学归纳法, $\forall b \in \mathbf{Z}^+$, $\|(W(k + b), H(b + 2)) - (W(k), H(k))\| \rightarrow 0$. \square

根据定理 4, $\forall \epsilon_1 > 0, \forall b \in \mathbf{Z}^+, \exists M(\epsilon_1, b) \in \mathbf{Z}^+, \|(W(k + b), H(b + 2)) - (W(k), H(k))\| < \epsilon_1$, 只要 $k > M(\epsilon_1, b)$; 或者 $\forall M \in \mathbf{Z}^+, \forall b \in \mathbf{Z}^+, \exists \epsilon_2(M, b) > 0, \|(W(k + b), H(b + 2)) - (W(k), H(k))\| < \epsilon_2(M, b)$, 只要 $k > M$. 也就是说, 经过足够步的迭代后, 再执行 b 步迭代, SENSEC 的输出不会被显著地改变; 或者说, 经过一定步数的迭代后, 再执行 b 步迭代, 导致的 SENSEC 输出的变化可被控制在一定范围内.

需要指出的是, 当迭代步长取值合理 (能使目标函数单调下降) 时, Hoyer 算法也具有上述稳定性, 但为 Hoyer 算法选定合理的迭代步长通常很难做到.

3.4 SENSE 的收敛性

定理 5 (SENESC 的收敛性定理). 定义解集合 $\Omega = \{\min_{0 \leq W, \epsilon \leq H; \forall i, \|W_{.i}\|_2=1} \|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij}$ 的稳定点}, SENSEC 关于 Ω 收敛.

证明. 依引理 6, $\{(W(k), H(k))\}$ 有界, 所以, $\{(W(k), H(k))\}$ 中存在收敛子列 $\{(W(k_l), H(k_l))\}$, 令 $\{(W(k_l), H(k_l))\} \rightarrow (\bar{W}, \bar{H})$.

根据式 (36), 固定 $\forall i, \{W_{.1}, W_{.2}, \dots, W_{.L}\} / W_{.i}$ 和 H 后, $\|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij}$ 关于 $W_{.i}$ 是严格凸的, 所以 $\forall W_{.i} \geq 0, \|V - \bar{W}\bar{H}\|_F^2 + 2\lambda \times \sum_{ij} H_{ij} \leq \|V - [\bar{W}_{.1}, \dots, \bar{W}_{.i-1}, W_{.i}, \bar{W}_{.i+1}, \dots, \bar{W}_{.L}]\bar{H}\|_F^2 + 2\lambda \sum_{ij} \bar{H}_{ij}$. 那么

$$\forall W_{.i} \geq 0, \left[\frac{\partial(\|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij})}{\partial W_{.i}} \Big|_{(W, H=(\bar{W}, \bar{H}))} \right]^T \times (W_{.i} - \bar{W}_{.i}) \geq 0, \quad \forall i \in \{1, 2, \dots, L\} \quad (44)$$

因为此时从点 $\bar{W}_{.i}$ 出发在区域 $\{W_{.i} | W_{.i} \geq 0\}$ 内没有下降方向. 同理,

$$\forall H_j. \geq \epsilon,$$

$$\left[\frac{\partial(\|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij})}{\partial H_j.} \Big|_{(W, H=(\bar{W}, \bar{H}))} \right]^T \times (H_j. - \bar{H}_j.) \geq 0, \quad \forall j \in \{1, 2, \dots, L\} \quad (45)$$

式 (45) 中, $\forall H_j. \geq \epsilon$ 而不是 $\forall H_j. \geq 0$ 的原因是构造 SENSEC 是基于推论 1 而非定理 1.

式 (44) 和式 (45) 合在一起构成了 (\bar{W}, \bar{H}) 为 $\min_{0 \leq W, \epsilon \leq H; \forall i, \|W_{.i}\|_2=1} \|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij}$ 的稳定点的充分条件^[15].

定义解集合 $\Omega = \{\min_{0 \leq W, \epsilon \leq H; \forall i, \|W_{.i}\|_2=1} \|V - WH\|_F^2 + 2\lambda \sum_{ij} H_{ij}$ 的稳定点}, $\{(W(k_l), H(k_l))\}$ 为任取的, 所以 $\{(W(k), H(k))\}$ 的所有极限点均在 Ω 中, 根据算法收敛的定义^[16-17], SENSEC 关于 Ω 收敛. \square

4 实验

本节用实验直观地分析 SENSEC 的性质. 实验内容分别针对迭代特性、解的质量、算法效率、描述能力与稀疏性调节因子 λ 的关系以及编码稀疏度与稀疏性调节因子 λ 的关系. 为便于与第 1 节列出的实验结果比较, 本节所有实验仍基于 ORL 库的 400 幅图像进行.

4.1 迭代特性

取 L 和 λ 均为 100, 图 9 记录了此时 SENSEC 的迭代曲线, 为便于比较它与 Hoyer 算法的区别, 图 9 中同时画出了当 $\mu = 1E-7, 1E-8$ 和 $1E-9$ 时的三条 Hoyer 算法迭代曲线. 很明显, Hoyer 算法的性能受迭代步长的影响很大, 只有选好迭代步长, 算法性能才能令人满意; 而 SENSEC 不涉及任何用户定义优化参数, 迭代过程中目标函数值单调下降, 因此它比 Hoyer 算法性能稳定、易于使用.

4.2 算法效率

取 L 和 λ 均为 100, 图 10 记录了 SENSEC 和 Hoyer 算法 (依图 9, 取 $\mu = 1E-9$) 的较详细迭代过程 (包括基于完整目标函数的迭代曲线和基于目标函数各部分性能的迭代曲线). 这里记录的迭代时长为 3000 s, 这是一个较长的运算时间, 因为这时目标函数已无明显下降趋势 (由图 10 (a) 可见), 因此由 3000 s 的迭代过程基本可分析算法的效率.

由图 10 (a) 可见, SENSEC 比 Hoyer 算法有快得多的下降速度, 经过 200 s ~ 300 s 的迭代, SENSEC 就可使目标函数值下降到与经过 3000 s 迭代时几乎相等的程度, 而 Hoyer 算法需要经过 2600 s 以上的迭代才能做到这点. 因此, SENSEC 的效率比 Hoyer 算法高很多, 这验证了第 3.2 节中的推断. 此外, 值得

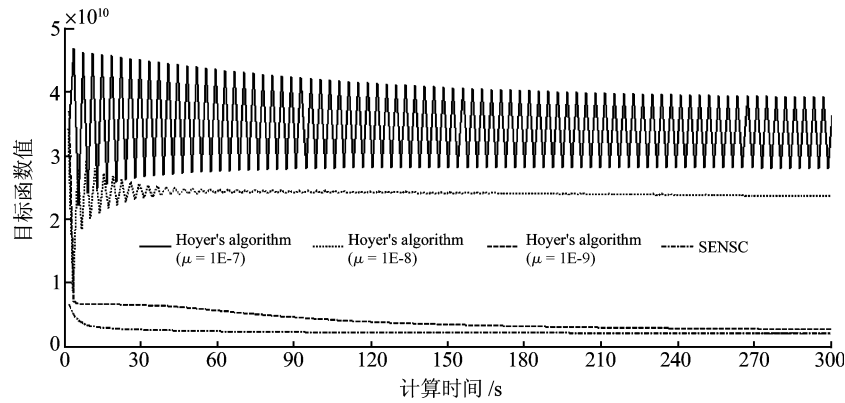


图 9 SENS 与 Hoyer 算法的迭代过程比较 ($L = 100, \lambda = 100$)

Fig. 9 Comparison of iterative processes between SENS and Hoyer's algorithm ($L = 100, \lambda = 100$)

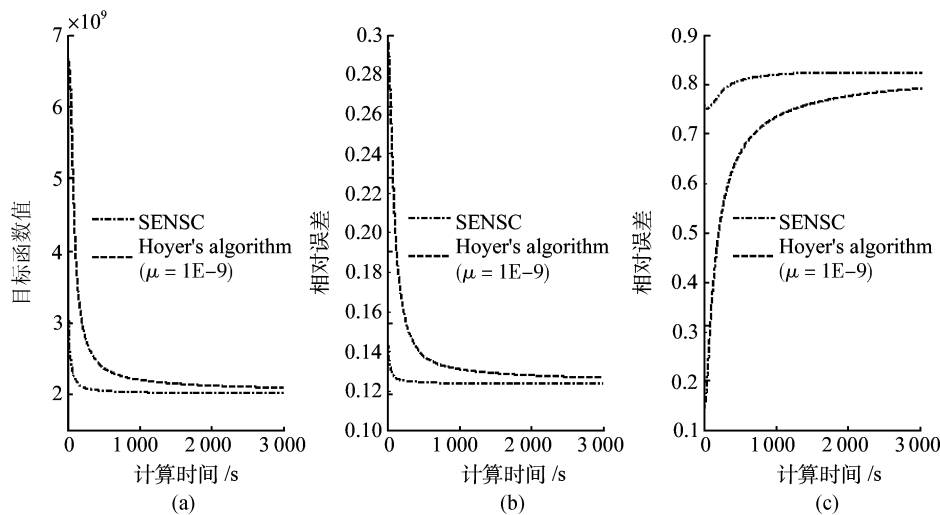


图 10 SENS 与 Hoyer 算法在解的质量和算法效率上的比较 ($L = 100, \lambda = 100$)

(a) 迭代结果由 NSC 目标函数值评价和描述; (b) 迭代结果由相对误差 (对应 NSC 目标函数的前半部分) 评价和描述; (c) 迭代结果由 NSC 编码稀疏度 (对应 NSC 目标函数的后半部分) 评价和描述

Fig. 10 Comparisons on solution qualities and algorithmic efficiencies between SENS and Hoyer's algorithm ($L = 100, \lambda = 100$)

(a) Iterative outcomes are represented and evaluated by NSC objective function value; (b) Iterative outcomes are represented and evaluated by relative error (the first part of NSC objective function); (c) Iterative outcomes are represented and evaluated by sparseness of NSC codes (the second part of NSC objective function)

注意的是, SENS 的高效率均衡地体现在对目标函数的每一部分 (描述的准确性和编码稀疏度的每一个) 的优化上. 由图 10 (b) 和图 10 (c) 知, 无论是就 NSC 结果对原始数据的相对误差而言, 还是就编码的稀疏度而言, SENS 均比 Hoyer 算法高效得多.

4.3 解的质量

如前所述, 迭代 3000s 时目标函数已无明显下降趋势, 此时解的质量基本反映了算法收敛 (或停步) 时解的质量. 由图 10 (a) 可见, SENS 能够比 Hoyer 算法收敛到更好的解, 因为 SENS 收敛处

的目标函数值比 Hoyer 算法收敛处的目标函数值小, 这验证了第 3.1 节中的推断. 并且, 图 10 (b) 说明 SENS 收敛处的相对误差值小于 Hoyer 算法收敛处的相对误差值, 图 10 (c) 说明 SENS 收敛处的编编码稀疏度远大于 Hoyer 算法收敛处的编编码稀疏度. 由此可知, 用目标函数的任一部分评价, “SENS 比 Hoyer 算法解更优” 这一结论均成立.

4.4 描述能力与稀疏性调节因子的关系

取 $L = 100$, 图 11 记录了用 SENS 进行 NSC 时, 相对误差值于不同 λ 值下在迭代过程中的变化.

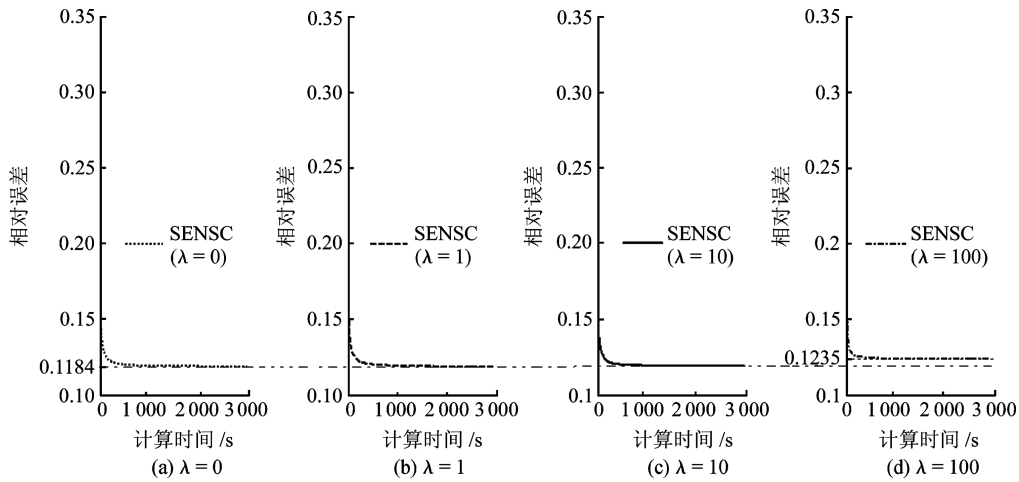


图 11 用 SENSIC 时, 描述能力与稀疏性调节因子 λ 的关系 ($L = 100$)

Fig. 11 Relationship between approximation accuracy and sparse factor λ while using SENSIC ($L = 100$)

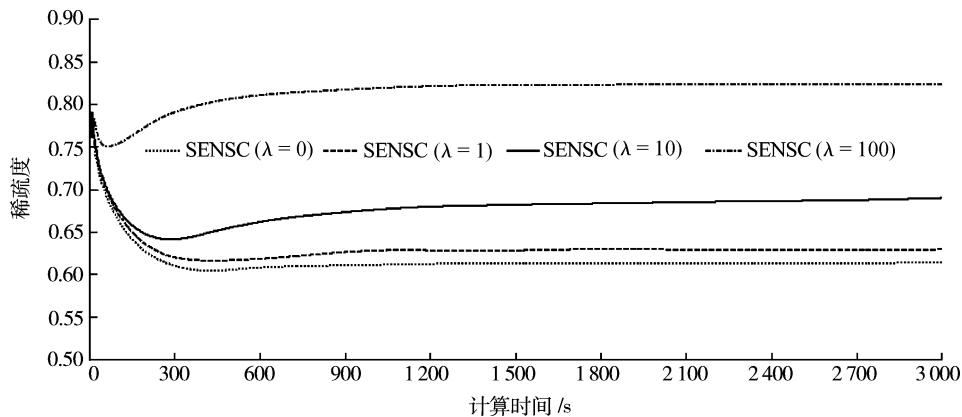


图 12 用 SENSIC 时, 码稀疏度与稀疏性调节因子 λ 的关系 ($L = 100$)

Fig. 12 Relationship between sparseness degree of codes and sparse factor λ while using SENSIC ($L = 100$)

$\lambda = 0$ 时, SENSIC 退化为基于最小二乘的 NMF 算法, 此时的实验结果如图 11 (a) 所示. 由对图 11 (b) ~ 11 (d) 中的实验记录观察可知, 当 λ 分别取 1 和 10 时, 由 SENSIC 进行的 NSC 对数据的描述能力不比由 SENSIC 进行的 NMF 差, 当 λ 取 100 时, 由 SENSIC 进行的 NSC 对数据的描述能力比由 SENSIC 进行的 NMF 稍差. 回顾图 4 可知, Hoyer 算法的描述能力受 λ 取值的影响情况与 SENSIC 相仿, 但就绝对数值而言, 每一个 λ 值下, 由 SENSIC 进行 NSC 算得的相对误差值, 均小于由 Hoyer 算法进行 NSC 算得的相对误差值 (见图 4 和图 11 (a) 上点划线对应的数值以及图 11 (d) 上虚线对应的数值). 回顾图 2 可知, 就描述能力而言, SENSIC 比 SNMF 算法强很多.

4.5 编码稀疏度与稀疏性调节因子 λ 的关系

取 $L = 100$, 图 12 记录了用 SENSIC 进行 NSC 时, 编码稀疏度于不同 λ 值下在迭代过程中的变化.

图 12 中对应 $\lambda = 0, 1, 10$ 和 100 的四条迭代曲线的散开性说明 SENSIC 控制编码稀疏性的能力很强. 比较图 5 和图 12 可知, SENSIC 对编码稀疏性的控制力好于 Hoyer 算法. 比较图 3 和图 12 可知, SENSIC 对编码稀疏度性的控制力远强于 SNMF 算法.

5 结论

NSC 是一个可应用于多学科研究的数据处理工具. Hoyer 算法和 SNMF 算法是目前仅有的两个进行 NSC 的算法. 应用中, 目前均使用 Hoyer 算法, 但它的性能受迭代步长影响很大 (不通过实验很难取到合适的迭代步长; 对不同数据, 不可能用同一迭代步长), 且效率较低. SNMF 算法无用户定义优化参数, 但它的性能远不如取合适迭代步长时的 Hoyer 算法. 为提高 NSC 的可应用性, 本文把 NSC 问题转化为对一组凸超抛物面函数做交替最小

化, 借助凸超抛物面的特性、点到非负点集的投影规则以及点到原点处单位超球的投影规则, 提出了一个新的 NSC 算法 — SENS. SENS 像 SNMF 算法那样, 无用户定义的优化参数, 因此其性能稳定. 理论推断和实验表明, SENS 比取合理迭代步长时的 Hoyer 算法性能更优, 这体现在: 1) 它的解好于 Hoyer 算法的解; 2) 它的效率高于 Hoyer 算法; 3) SENS 调节编码稀疏性的能力强于 Hoyer 算法. 此外, 文中证明了 SENS 的稳定性和收敛性.

References

- Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, **381**(6583): 607–609
- Hoyer P O. Non-negative sparse coding. In: Proceedings of IEEE Workshop on Neural Networks for Signal Processing. Martigny, Switzerland: IEEE, 2002. 557–565
- Hoyer P O. Modeling receptive fields with non-negative sparse coding. *Neurocomputing*, 2003, **52-54**: 547–552
- Palmer S E. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, **9**(4): 441–474
- Wachsmuth E, Oram M W, Perrett D I. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, 1994, **4**(5): 509–522
- Logothetis N K, Sheinberg D L. Visual object recognition. *Annual Review of Neuroscience*, 1996, **19**(1): 577–621
- Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, **401**(6755): 788–791
- Hoyer P O, Hyvarinen A. A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*, 2002, **42**(12): 1593–1605
- Shang Li, Huang De-Shuang, Zheng Chun-Hou. Natural image denoising method based on non-negative sparse coding shrinkage. *Journal of University of Science and Technology of China*, 2006, **36**(5): 497–501
(尚丽, 黄德双, 郑春厚. 非负稀疏编码收缩法的自然图像消噪. 中国科学技术大学学报, 2006, **36**(5): 497–501)
- Shastri B J, Levine M D. Face recognition using localized features based on non-negative sparse coding. *Machine Vision and Applications*, 2007, **18**(2): 107–122
- Liu W X, Zheng N N, Lu X F. Non-negative matrix factorization for visual coding. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Hong Kong, China: IEEE, 2003. 293–296
- Lee D D, Seung H S. Algorithms for non-negative matrix factorization. *Advance in Neural Information Processing Systems*. Cambridge: MIT Press, 2000. 556–562
- AT & T Laboratories Cambridge. The database of face [Online], available: <http://www.cl.cam.ac.uk/research/dtg/archive/facedatabase.html>, April 8, 2008
- Hoyer P O. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 2004, **5**(9): 1457–1469
- Bertsekas D P. *Nonlinear Programming (Second Edition)*. Massachusetts: Athena Scientific, 1999
- Luenberger D G. *Introduction to Linear and Nonlinear Programming*. Massachusetts: Addison-Wesley, 1973
- Chen Bao-Lin. *Theories and Algorithms of Optimization*. Beijing: Tsinghua University Press, 1989
(陈宝林. 最优化理论与算法. 北京: 清华大学出版社, 1989)



李 乐 讲师, 博士. 主要研究方向为非负矩阵分解、非负矩阵集分解及其在图像工程中的应用. 本文通信作者.

E-mail: lile05@mails.tsinghua.edu.cn

(LI Le Lecturer and Ph.D. at Tsinghua University. His research interest covers nonnegative matrix factorization, nonnegative matrix-set factorization and their applications on image engineering. Corresponding author of this paper.)



章毓晋 清华大学电子工程系教授. 主要研究方向为图像工程 (图像处理、图像分析、图像理解及其技术应用).

E-mail: zhang-yj@tsinghua.edu.cn

(ZHANG Yu-Jin Professor at Tsinghua University. His research interest covers image engineering that includes image processing, image analysis and image understanding, as well as their applications.)