DOI:10.13878/j.cnki.jnuist.2020.03.015

陶盈吟[1]　杨仪[1]　代祥光[1]　苏晓杰[2]

基于 KKT 条件的稀疏编码算法收敛性研究

摘要

本文提出了基于 KKT 条件的稀疏编码算法.首先,将非凸非光滑的稀疏编码问题分解成两个凸非光滑问题;然后,巧妙地运用两个矩阵使两个凸非光滑问题转换成三个光滑凸优化问题,并通过 KKT 条件对三个问题进行求解,再通过凸优化理论证明三个问题在其对应规则下是非增的.最后,实验结果验证了算法的收敛性.

关键词

KKT 条件;收敛性;非凸非光滑;稀疏编码

中图分类号 O232

文献标志码 A

0　引言

稀疏编码是一种无监督学习方法,即从输入样本数据中学习一系列基,这些基非常符合人类的视觉感知[1-2].当基的维度远高于原始数据的维度,那么稀疏编码可以理解为超完备基学习方法;当基的维度远低于输入样本的维度,稀疏编码可以理解为数据降维学习方法.由于稀疏编码的多重特性,它经常被应用于模式识别、聚类和信号处理[3-8]等.

假设输入样本为一个向量 $\boldsymbol{\xi} \in \mathbf{R}^m$,稀疏编码需要找到基向量 $\boldsymbol{b}_1,\cdots,\boldsymbol{b}_r \in \mathbf{R}^m$ 和权重向量 $\boldsymbol{s} \in \mathbf{R}^r$ 使得:

$$\boldsymbol{\xi} \approx \sum_{j=1}^{r} \boldsymbol{b}_j \boldsymbol{s}_j, \tag{1}$$

其中,r 是分解因子.若输入样本为多个向量 $\boldsymbol{X}=[\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_n]$,则需要找到基向量 $\boldsymbol{B}=[\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n]_r \in \mathbf{R}^{m\times r}$ 和权重向量 $\boldsymbol{S}=[\boldsymbol{s}_1,\cdots,\boldsymbol{s}_n] \in \mathbf{R}^{r\times n}$ 使得:

$$\boldsymbol{X} \approx \boldsymbol{BS}. \tag{2}$$

一般地,采用欧几里得距离测量 $\boldsymbol{X}$ 和 $\boldsymbol{BS}$,式(2)可以转换成如下优化问题:

$$\min_{B,S} F(\boldsymbol{B},\boldsymbol{S}) = \| \boldsymbol{X}-\boldsymbol{BS} \|_F^2 + 2\lambda \| \boldsymbol{S} \|_1, \tag{3}$$

其中,$\| \boldsymbol{X} \|_F = \sqrt{\sum_{i,j} x_{ij}^2}$ 和 $\| \boldsymbol{X} \|_1 = \sum_{i,j} | x_{ij} |$.式(3)中的第1项用于测量输入样本矩阵和近似矩阵之间的误差,第 2 项为稀疏项,λ 越大说明 $\boldsymbol{S}$ 越稀疏,反之亦然.

稀疏编码问题可以被看作为非凸优化非光滑问题,无法直接用传统的优化算法直接求解.常见的交替优化框架[9]将问题(3)转换成两个非凸优化问题,然后使用传统优化算法交替求解直至收敛.简单来说,固定矩阵 $\boldsymbol{B}$,优化 $\boldsymbol{S}$,然后固定 $\boldsymbol{S}$,优化 $\boldsymbol{B}$.假设,第 k 次解 $\boldsymbol{B}^k$ 和 $\boldsymbol{S}^k$ 已经得到,那么 $\boldsymbol{B}$ 和 $\boldsymbol{S}$ 的局部最优解可以通过求解以下两个式子得到:

$$\boldsymbol{B}^{k+1} = \arg\min_B F(\boldsymbol{B}) = \| \boldsymbol{X}-\boldsymbol{BS}^k \|_F^2, \tag{4}$$

$$\boldsymbol{S}^{k+1} = \arg\min_S F(\boldsymbol{S}) = \| \boldsymbol{X}-\boldsymbol{B}^{k+1}\boldsymbol{S} \|_F^2 + 2\lambda \| \boldsymbol{S} \|_1. \tag{5}$$

本文提出采用 KKT 条件和梯度优化法来优化问题(3),主要贡献如下:1)巧妙地运用两个非负矩阵代替矩阵 $\boldsymbol{S}$,将稀疏编码问题转换成非负约束的矩阵分解问题;2)建立对应的拉格朗日函数,并采用

收稿日期 2020-02-26

资助项目 重庆市高校市级重点实验室资助项目([2017]3);重庆市发展和改革委员会资助项目(2017[1007]);重庆市教委科技研究项目(KJQN201901203,KJQN201901218,KJ1710248);重庆市自然科学基金(cstc2019jcyj-bshX0101)

作者简介

陶盈吟,女,硕士生,研究方向为优化算法、机器学习.597370352@qq.com

杨仪(通信作者),女,博士,研究方向为神经网络、非线性动力系统. yang1595@126.com

1　重庆三峡学院　智能信息处理与控制重庆高校市级重点实验室,重庆,404100

2　重庆大学　自动化学院,重庆,400044

KKT 条件设计其迭代规则;3)通过优化理论证明迭代规则的收敛性.

1 稀疏编码迭代规则

问题(3)是非凸非光滑问题,无法搜索到全局最优解.本节提出基于 KKT 条件和梯度下降法的迭代规则来搜索问题(3)的局部最优解.

首先,给定两个非负矩阵 $\boldsymbol{U}=\max(\boldsymbol{S},0)$ 和 $\boldsymbol{V}=\max(-\boldsymbol{S},0)$,那么

$$\boldsymbol{S}=\boldsymbol{U}-\boldsymbol{V}. \tag{6}$$

将式(6)代入问题(3),那么问题(3)变成非凸光滑问题:

$$\begin{aligned}&\min_{B,U,V} F(\boldsymbol{B},\boldsymbol{U},\boldsymbol{V})=\\&\quad \|\boldsymbol{X}-\boldsymbol{BU}+\boldsymbol{BV}\|_F^2+2\lambda\|\boldsymbol{U}\|_1+2\lambda\|\boldsymbol{V}\|_1\\&\text{s.t. } \boldsymbol{U}\geqslant 0,\boldsymbol{V}\geqslant 0.\end{aligned} \tag{7}$$

将(7)的目标函数 $F(\boldsymbol{B},\boldsymbol{U},\boldsymbol{V})$ 变换如下:

$$\begin{aligned}F(\boldsymbol{B},\boldsymbol{U},\boldsymbol{V})=&\operatorname{tr}((\boldsymbol{X}-\boldsymbol{BU}+\boldsymbol{BV})^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{BU}+\\&\boldsymbol{BV}))+2\lambda\sum_{ij}u_{ij}+2\lambda\sum_{ij}v_{ij}=\\&\operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})-2\operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{BU})+2\operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{BV})-\\&2\operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})+\operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU})+\\&\operatorname{tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})+2\lambda\sum_{ij}u_{ij}+2\lambda\sum_{ij}v_{ij},\end{aligned}$$

其中,$\operatorname{tr}(\cdot)$ 为矩阵的迹.

其次,给出问题(7)的拉格朗日函数如下:

$$\begin{aligned}L(\boldsymbol{B},\boldsymbol{U},\boldsymbol{V})=&\operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})-2\operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{BU})+\\&2\operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{BV})-2\operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})+\\&\operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU})+\operatorname{tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})+\\&2\lambda\sum_{ij}u_{ij}+2\lambda\sum_{ij}v_{ij}+\operatorname{tr}(\boldsymbol{\Phi U})+\operatorname{tr}(\boldsymbol{\Psi V}),\end{aligned} \tag{8}$$

其中,$\boldsymbol{\Phi}=[\phi_{ij}]$ 和 $\boldsymbol{\Psi}=[\psi_{ij}]$ 是约束项 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的拉格朗日乘子项.那么对 $L(\boldsymbol{B},\boldsymbol{U},\boldsymbol{V})$ 对 $\boldsymbol{B},\boldsymbol{U},\boldsymbol{V}$ 的偏导如下:

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{B}}=-2\boldsymbol{X}(\boldsymbol{U}-\boldsymbol{V})^{\mathrm{T}}+2\boldsymbol{B}(\boldsymbol{U}-\boldsymbol{V})(\boldsymbol{U}-\boldsymbol{V})^{\mathrm{T}}, \tag{9}$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{U}}=-2\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{X}+\boldsymbol{BV})+2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU}+2\lambda\boldsymbol{I}+\boldsymbol{\Phi}, \tag{10}$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{V}}=2\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{BU})+2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV}+2\lambda\boldsymbol{I}+\boldsymbol{\Psi}, \tag{11}$$

其中,矩阵 $\boldsymbol{I}$ 是元素全1的矩阵.问题(9)是无约束的凸优化问题,牛顿法得到最优解如下:

$$\boldsymbol{B}\leftarrow\boldsymbol{B}-\boldsymbol{H}(\boldsymbol{B})^{-1}\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{B}}, \tag{12}$$

其中海塞矩阵 $H(\boldsymbol{B})=\dfrac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{B}^2}=2(\boldsymbol{U}-\boldsymbol{V})(\boldsymbol{U}-\boldsymbol{V})^{\mathrm{T}}$.若输入样本矩阵维度过高,海塞矩阵求逆过于复杂,采用拟牛顿或 BFGS 算法更好.基于(8)的 KKT 条件 $\phi_{ij}u_{ij}=0$ 和 $\psi_{ij}v_{ij}=0$,那么得到:

$$(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij}u_{ij}+(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})_{ij}u_{ij}-(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU})_{ij}u_{ij}-(\lambda\boldsymbol{I})_{ij}u_{ij}=0, \tag{13}$$

$$(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU})_{ij}v_{ij}-(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij}v_{ij}-(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})_{ij}v_{ij}-(\lambda\boldsymbol{I})_{ij}v_{ij}=0. \tag{14}$$

等式(13)和(14)可以得到 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的最优解如下:

$$u_{ij}\leftarrow u_{ij}\frac{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})_{ij}}{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU})_{ij}-(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij}+(\lambda\boldsymbol{I})_{ij}}, \tag{15}$$

$$v_{ij}\leftarrow v_{ij}\frac{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU})_{ij}}{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BV})_{ij}+(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij}+(\lambda\boldsymbol{I})_{ij}}. \tag{16}$$

2 收敛性证明

问题(7)的局部最优解可以根据迭代规则(12),(15)和(16)得到,那么,(12),(15)和(16)同样是以下三个问题的全局最优解:

$$\min_{B} F(\boldsymbol{B})=\|\boldsymbol{X}-\boldsymbol{BU}+\boldsymbol{BV}\|_F^2, \tag{17}$$

$$\min_{U\geqslant 0} F(\boldsymbol{U})=\|\boldsymbol{X}-\boldsymbol{BU}+\boldsymbol{BV}\|_F^2+2\lambda\|\boldsymbol{U}\|_1, \tag{18}$$

$$\min_{V\geqslant 0} F(\boldsymbol{V})=\|\boldsymbol{X}-\boldsymbol{BU}+\boldsymbol{BV}\|_F^2+2\lambda\|\boldsymbol{V}\|_1. \tag{19}$$

定义 1 若 $G(\boldsymbol{X},\boldsymbol{X}')$ 是目标函数 $F(\boldsymbol{X})$ 的辅助函数,$G(\boldsymbol{X},\boldsymbol{X}')$ 必须满足以下条件:

$$G(\boldsymbol{X},\boldsymbol{X}')\geqslant F(\boldsymbol{X}),G(\boldsymbol{X},\boldsymbol{X})=F(\boldsymbol{X}). \tag{17}$$

引理 1 如果 $G(\boldsymbol{X},\boldsymbol{X}')$ 是目标函数 $F(\boldsymbol{X})$ 辅助函数,那么 $F(\boldsymbol{X})$ 在以下迭代规则是非增的:

$$x^{t+1}=\arg\min G(\boldsymbol{x},\boldsymbol{x}^t), \tag{18}$$

其中 X^t 是 $F(X)$ 的第 t 次解.

引理 2 定义函数

$$\begin{aligned}G(u,u_{ab}^t)=&F_{ab}(u_{ab}^t)+F'_{ab}(u_{ab}^t)(u-u_{ab}^t)+\\&\frac{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{BU})_{ij}+(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij}+(\lambda\boldsymbol{I})_{ij}}{u_{ab}^t}(u-u_{ab}^t)^2\end{aligned} \tag{19}$$

是 $F(\boldsymbol{U})$ 的辅助函数.

证明 根据定义1,$G(u,u)=F_{ab}(u)$ 明显成立.只需证明 $G(u,u_{ab}^t)\geqslant F_{ab}(u_{ab}^t)$,即可得到 $G(u,u_{ab}^t)$ 是 $F(\boldsymbol{U})$ 的辅助函数.首先,将 $F_{ab}(u)$ 在 u_{ab}^t 展开,得到以下式子

$$\begin{aligned}F_{ab}(u)=&F_{ab}(u_{ab}^t)+F'_{ab}(u_{ab}^t)(u-u_{ab}^t)+\\&(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})_{aa}(u-u_{ab}^t)^2,\end{aligned} \tag{20}$$

那么,

$$G(u,u_{ab}^t) - F_{ab}(u_{ab}^t) = \left(\frac{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{U})_{ij} - (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij} + (\lambda\boldsymbol{I})_{ij}}{u_{ab}^t} - (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})_{bb}\right)(u-u_{ab}^t)^2. \tag{21}$$

显然, $(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{U})_{ab} = \sum_{l=1}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})_{al}u_{lj}^t\, b \geqslant (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})_{aa}$ 以及 $(-(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij}+(\lambda\boldsymbol{I})_{ij})u_{ab}^t \geqslant 0$.因此, $G(u,u_{ab}^t) \geqslant F_{ab}(u_{ab}^t)$.

引理 3 定义函数

$$K(v,v_{ab}^t) = F_{ab}(v_{ab}^t) + F'_{ab}(v_{ab}^t)(v-v_{ab}^t) + \frac{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{V})_{ij} + (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij} + (\lambda\boldsymbol{I})_{ij}}{v_{ab}^t}(v-v_{ab}^t)^2 \tag{22}$$

是 $F(\boldsymbol{V})$ 的辅助函数.证明详见引理 2.

定理 1 $F(\boldsymbol{U})$ 和 $F(\boldsymbol{V})$ 在迭代规则(15)和(16)下是单调非增的.

证明 根据引理 1 和 2 以及 $F(\boldsymbol{U})$,我们可以得到

$$u_{ab}^t = u_{ab}^t - u_{ab}^t\frac{F'_{ab}(u_{ab}^t)}{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{U})_{ij} + (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij} + (\lambda\boldsymbol{I})_{ij}} = u_{ab}^t\frac{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{V})_{ij}}{(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{U})_{ij} + (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X})_{ij} + (\lambda\boldsymbol{I})_{ij}}, \tag{23}$$

那么, $F(\boldsymbol{U})$ 在迭代规则(15)下是单调非增的.同理,$F(\boldsymbol{V})$ 在迭代规则(16)下是单调非增的.

3 实验结果

本文算法(KKTSC)将与算法 Feature-sgin[10] 和 MexLasso[11] 在收敛性和基学习做分析比较.本文使用人脸数据集 Olivetti Research Laboratory(ORL)来验证收敛性和基学习有效性.ORL 人脸数据集由英国剑桥 Olivetti 实验室拍摄的一系列人脸图像组成,共计 40 个不同性别、种族和年龄的对象.每个对象有 10 幅像素为 92×112 的灰度图像.人脸细节和表情均有不同,比如睁眼或闭眼、是否戴眼镜、笑或不笑、有无淡光、不同的人脸姿态和人脸尺寸等.

三种算法将在 ORL 上做收敛性分析比较.选择分解因子 r 分别为 100,500,1 000,其中收敛性比较结果如表 1 所示.当分解因子 r 较小时(比如 r 为 100,500),三种算法都可以在有限时间得到目标函数值,MexLasso 可以在短时间获得目标函数值.当分解因子 $r=1\,000$,仅 KKTSC 和 MexLasso 可以在有限时间得到目标函数值,Feature-sign 无法在有限时间获得目标函数值.综上所述,KKTSC 随着数据维度增加,其收敛速度比另两种算法更快.

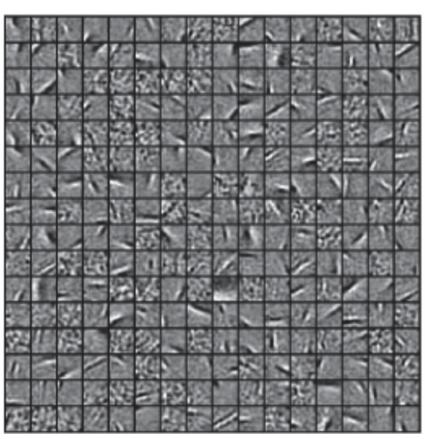
表 1 KKTSC、Feature-sgin 和 MexLasso 之间的收敛性结果

Table 1 Convergence results of KKTSC, Feature-sgin, and MexLasso

算法	收敛时间			目标函数值		
	$r=100$	$r=500$	$r=1\,000$	$r=100$	$r=500$	$r=1\,000$
KKTSC	5.821	80.337	260.126	129 931	96 544	57 259
Feature-sign	1.348	58.258	536.010	129 931	96 544	57 259
MexLasso	28.821	1 039.233	—	130 102	96 546	—

由于 MexLasso 在高维数据中收敛性较慢,这里只考虑将 KKTSC 和 Feature-sign 做超完备基学习.设置迭代次数为 50 次,图 1 给出了完备基学习结果.根据图 1,两种算法都可以完成超完备基的学习.然而,KKTSC 只需要 24 min 完成超完备基的学习,而 Feature-sign 则需要 58 min 完成超完备基的学习.

a. KKTSC(50次迭代, 24 min)



b. Feature-sign(50次迭代, 58 min)

图 1 KKTSC 和 Feature-sign 之间的超完备基学习结果

Fig. 1 The over-complete bases learning result of KKTSC (a) and Feature-sign(b)

4 结束语

本文将稀疏编码问题分解成三个凸优化问题,然后通过 KKT 给定其迭代规则.通过凸优化理论证明三个凸优化问题在其对应规则下是非增的.最后,通过 ORL 数据集验证了所提出算法具有快速学习超完备基的能力.

参考文献

References

[1] Olshausen B A,Field D J.Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J].Nature,1996,381(6583):607-609
[2] Olshausen B A,Field D J.Sparse coding with an overcomplete basis set:a strategy employed by *V*1? [J].Vision Research,1997,37(23):3311-3325
[3] Labusch K,Barth E,Martinetz T.Simple method for high-performance digit recognition based on sparse coding[J].IEEE Trans Neural Netw,2008,19(11):1985-1989
[4] He B,Xu D,Rui N,et al.Fast face recognition via sparse coding and extreme learning machine[J].Cognitive Computation,2014,6(2):264-277
[5] Zheng M,Bu J J,Chen C,et al.Graph regularized sparse coding for image representation[J].IEEE Transactions on Image Processing:a Publication of the IEEE Signal Processing Society,2011,20(5):1327-1336
[6] Shu Z Q,Zhao C X,Huang P.Constrained sparse concept coding algorithm with application to image representation[J].KSII Transactions on Internet and Information Systems,2014,8(9):3211-3230
[7] Lörincz A,Palotai Z,Szirtes G.Efficient sparse coding in early sensory processing:lessons from signal recovery[J].PLoS Computational Biology,2012,8(3):e1002372
[8] Zhao Y X,Liu Z Y,Wang Y Y,et al.Sparse coding algorithm with negentropy and weighted 1-norm for signal reconstruction[J].Entropy,2017,19(11):599
[9] Bertsekas D P.Nonlinear programming[J].Journal of the Operational Research Society,1997,48:332-334
[10] Lee H,Battle A,Raina R,et al.Efficient sparse coding algorithms[J].Proc of Nips,2007,19:801-808
[11] Mairal J,Bach F,Ponce J,et al.Online learning for matrix factorization and sparse coding[J].Journal of Machine Learning Research,2009,11:19-60

Convergence of sparse coding based on KKT conditions

TAO Yingyin[1] YANG Yi[1] DAI Xiangguang[1] SU Xiaojie[2]
1 Key Laboratory of Intelligent Information Processing and Control of Chongqing Municipal Institutions of Higher Education,Chongqing Three Gorges University,Chongqing 404100
2 College of Automation,Chongqing University,Chongqing 400044

Abstract This paper proposes a sparse coding algorithm based on KKT conditions.Firstly,the non-convex non-smooth sparse coding problem is decomposed into two convex non-smooth problems.Secondly,the two convex non-smooth problems are skillfully transformed into three smooth convex optimization problems by using two matrices.Finally,the three problems are solved by KKT conditions.In addition,we prove the convergence of the algorithm.Meanwhile,experimental simulation shows the convergence of the algorithm.

Key words KKT conditions;convergence;non-convex non-mooth;sparse coding