http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0462

基于正则化秩 k 矩阵逼近的稀疏主成分分析

杨茜, 刘红英*

(北京航空航天大学 数学与系统科学学院, 北京 100083)

摘 要: 在计算稀疏主成分(PCs)时,由于同时求 k 个主成分的做法可以减少计算所产生的累积误差,因此提出了基于正则化秩 k 矩阵逼近的稀疏主成分模型,并设计了求解该模型的块坐标下降法(BCD-sPCA-rSVD)。该算法的主要思想是先把变量按坐标分成 $2k$ 个块,当固定其他 $2k-1$ 个坐标块的变量时,求解关于单个坐标块的子问题并给出子问题的显式解,循环地求解这些子问题直至满足终止条件。该算法每次迭代的计算复杂度关于样本个数与变量维数都是线性的,并且证明了它是收敛的。该算法不仅易于实现,数值仿真结果表明,该算法应用到真实数据与合成数据上都是可行且有效的。它不仅使累积误差降低,而且具有较低的计算复杂度,因而可以有效地求解大规模稀疏主成分分析问题。

关 键 词: 降维; 稀疏主成分; 正则化; 块坐标下降法; 奇异值分解; 阈值
中图分类号: O212.4; TP181; O221.2
文献标识码: A **文章编号**: 1001-5965(2017)06-1239-08

主成分分析(Principal Component Analysis, PCA)是一种重要的数据分析与降维工具,它在科学与工程领域都有着广泛的应用。然而,由于传统 PCA 中的主成分(Principal Component, PC)是所有原始变量的线性组合,即 PC 载荷一般都是非零的[1]。因此,很难对与主成分相关的原始因素进行解释。而稀疏主成分分析就是为了解决主成分的这一缺陷而提出的,它使得 PC 载荷中零元素的个数变多,使主成分的可解释性增强。

当需要计算 $k(>1)$ 个稀疏主成分时,主要有 2 种做法:第 1 种做法是依次求单个主成分,即求解“单个稀疏主成分”模型,并通过压缩的(deflated)数据矩阵或数据协方差阵来依次获得更多的稀疏主成分,包括基于 LASSO(Least Absolute Shrinkage and Selection Operator)约束的投影梯度法(SCoTLASS)[2]、稀疏主成分分析的半定规划直接表述法(DSPCA)[3]、基于正则化奇异值分解模型的稀疏主成分分析(sPCA-rSVD)[4]、贪心法[5-6]、单位步长条件梯度法(ConGradU)[7]、拉格朗日对偶法[8]和期望最大化稀疏主成分分析(EMPCA)[9]等。第 2 种做法是同时求多个主成分,即建立合适的优化模型一次获得多个稀疏主成分,包括稀疏主成分分析法(SPCA)[1]、广义幂法(Gpower $l_{1,k}$, Gpower $l_{0,k}$)[10]、稀疏主成分分析的增广拉格朗日函数法(ALSPCA)[11]和块坐标下降稀疏主成分分析法(BCD-SPCA)[12]等。虽然在给定不同的稀疏性参数下,依次求解单个主成分较为容易,但这种计算方式会产生累积误差,而同时计算多个主成分则减少了累积误差,并且这类做法对解决大规模稀疏 PCA 问题更有效且实用[12]。因而研究同时求多个稀疏主成分的算法具有应用价值。

本文提出了基于 l_1 正则化秩 k 矩阵逼近的稀疏 PCA 模型,并设计了求解该模型的块坐标下

收稿日期: 2016-05-31; **录用日期**: 2016-09-02; **网络出版时间**: 2016-10-10 09:03
网络出版地址: www.cnki.net/kcms/detail/11.2625.V.20161010.0903.004.html
基金项目: 国家自然科学基金 (61172060, 61403011)
* **通讯作者**: E-mail: liuhongying@buaa.edu.cn

引用格式: *杨茜, 刘红英. 基于正则化秩 k 矩阵逼近的稀疏主成分分析[J]. 北京航空航天大学学报, 2017, 43(6): 1239-1246. YANG Q, LIU H Y. Sparse principal component analysis via regularized rank-k matrix approximation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(6): 1239-1246 (in Chinese).*

降法(BCD-sPCA-rSVD)。该算法的主要思想是利用块坐标下降法把变量按坐标分成 $2k$ 个块,当固定其他 $2k-1$ 个坐标块的变量时,求解关于单个坐标块的子问题并给出子问题的显式解,循环地求解这些子问题直至满足终止条件。同时本文还分析了 BCD-sPCA-rSVD 的计算复杂度并证明其是收敛的。最后的数值仿真结果表明了该算法的可行性与有效性。

本文假设所有的向量都是列向量。这里列出文中常用的记号与术语。

记号 $\mathbf{R}^{n\times p}$ 和 sgn(·)分别为 $n\times p$ 维实矩阵和逐分量符号算子。$\boldsymbol{v}$ 为向量,它的第 i 个分量为 v_i。$\boldsymbol{V}$ 为矩阵,它的第 i 列用向量 $\boldsymbol{v}_i$ 表示。符号 $\|\boldsymbol{v}\|_2=\sqrt{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}$ 和 $\|\boldsymbol{v}\|_1=\sum_i|v_i|$ 分别为向量 $\boldsymbol{v}$ 的 l_2 范数和 l_1 范数,$\|\boldsymbol{v}\|_0$ 为 $\boldsymbol{v}$ 的 l_0 范数,它表示 $\boldsymbol{v}$ 中非零元素的个数。符号 $|\boldsymbol{v}|$ 和 $\boldsymbol{v}_+$ 是指其第 i 个分量分别为 $|v_i|$ 和 $\max(v_i,0)$ 的向量。$\boldsymbol{e}$ 为分量全为 1 的向量。记号 $S_{\gamma e}(\boldsymbol{v})=\mathrm{sgn}(\boldsymbol{v})(|\boldsymbol{v}|-\gamma\boldsymbol{e})_+$ 为向量 $\boldsymbol{v}$ 的软阈值算子。本文中,$\boldsymbol{X}\in\mathbf{R}^{n\times p}$ 为数据矩阵($\boldsymbol{X}$ 已作中心化处理,每列的平均值为 0),$\boldsymbol{W}=(\boldsymbol{w}_1,\boldsymbol{w}_2\cdots,\boldsymbol{w}_k)\in\mathbf{R}^{p\times k}$,$\boldsymbol{U}=(\boldsymbol{u}_1,\boldsymbol{u}_2\cdots,\boldsymbol{u}_k)\in\mathbf{R}^{n\times k}$,$\boldsymbol{V}=(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k)\in\mathbf{R}^{p\times k}$,其中 n 为样本个数,p 为变量个数,k 为主成分个数。符号 $\|\cdot\|_{\mathrm{F}}$ 表示矩阵的 Frobenius 范数。

1 相关的稀疏 PCA 模型与算法

本节主要对与 BCD-sPCA-rSVD 密切相关的 3 种稀疏 PCA 算法 SPCA、sPCA-rSVD 与 BCD-SPCA进行简单的回顾与说明。

1.1 SPCA 与 sPCA-rSVD 的模型与算法

Zou 和 Hastie[1] 在 2006 年提出的 SPCA,它的模型是先将 PCA 表述成回归型的优化问题,并添加关于回归系数的 LASSO[18](即 l_1 范数)惩罚项,即

$$\begin{aligned}&\min_{\boldsymbol{W},\boldsymbol{V}}\ \|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{V}\boldsymbol{W}^{\mathrm{T}}\|_{\mathrm{F}}^2+\lambda\sum_{j=1}^{k}\|\boldsymbol{v}_j\|_2^2+\sum_{j=1}^{k}\gamma_j\|\boldsymbol{v}_j\|_1\\&\text{s.t.}\quad\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{I}_k\end{aligned}\tag{1}$$

其中:正参数 $\gamma_j(j=1,2,\cdots,k)$ 用来控制载荷 $\boldsymbol{v}_j$ 的稀疏度;$\boldsymbol{I}_k$ 为 k 阶单位矩阵。对于 $n<p$ 的数据集,要求 l_2 范数前的参数 $\lambda>0$;当 $n>p$ 时,参数 λ 的设置是不必要的,取 $\lambda=0$。

对于模型(1),SPCA 利用块坐标下降法将模型(1)的变量分成 2 个坐标块,即 $\boldsymbol{W}$ 和 $\boldsymbol{V}$,当固定其中一个坐标块 $\boldsymbol{W}$(或 $\boldsymbol{V}$)的变量时,求解关于另外一个坐标块 $\boldsymbol{V}$(或 $\boldsymbol{W}$)的子问题,而求解关于 $\boldsymbol{V}$ 的子问题又可转化为求解 k 个独立的关于 $\boldsymbol{V}$ 的列向量 $\boldsymbol{v}_j$ 的子问题,交替地求解关于 $\boldsymbol{V}$ 和 $\boldsymbol{W}$ 的子问题直至终止,最后将 $\boldsymbol{v}_j(j=1,2,\cdots,k)$ 单位化可得到稀疏的 PC 载荷。SPCA 的优点是它的模型可将 PCA 与回归型问题建立联系,即求解去掉 LASSO 惩罚项的模型(1)就可得到主成分,并且对于不同类型的数据集($n>p$ 与 $n<p$),SPCA 都具有较低的计算复杂度[1]。

Shen 和 Huang[4] 在 2008 年提出了 sPCA-rSVD,此模型首先利用了数据矩阵奇异值分解与 PCA 的联系,由秩 1 矩阵逼近得到第一主成分,然后在对应的目标函数中添加正则化项(l_1、l_0 范数等)来得到稀疏的载荷,即

$$\begin{aligned}&\min_{\boldsymbol{u},\boldsymbol{v}}\ \|\boldsymbol{X}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\|_{\mathrm{F}}^2+P_\gamma(\boldsymbol{v})\\&\text{s.t.}\quad\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1\end{aligned}\tag{2}$$

其中:正则化项 $P_\gamma(\boldsymbol{v})$ 为 l_1 或 l_0 范数等,γ 为控制载荷 $\boldsymbol{v}$ 的稀疏度参数。一般来说,要求载荷 $\boldsymbol{v}$ 是单位向量,然而这一约束使得求解目标中含有关于 $\boldsymbol{v}$ 的惩罚项的模型变得困难,因此模型(2)是对 $\boldsymbol{v}$ 和 $\boldsymbol{u}$ 的一个尺度变换,使 $\boldsymbol{u}$ 为单位向量而 $\boldsymbol{v}$ 没有任何约束限制。

sPCA-rSVD 也是利用块坐标下降法将模型(2)的变量分成 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 这 2 个坐标块,当固定其中一个坐标块的变量时,求解关于另外一个坐标块的子问题,交替地求解关于这 2 个坐标块的子问题直至终止。当正则化项为 l_1 范数时(sPCA-rSVD-soft),sPCA-rSVD 中子问题的求解较为简单(主要的计算任务包括简单的线性回归和逐分量阈值),但是 sPCA-rSVD 依次求解主成分的计算方式会产生累积误差。

1.2 BCD-SPCA 的模型与算法

Zhao 等[12] 在 2015 年提出了 BCD-SPCA,它考虑了含有 PC 载荷的 l_1 或 l_0 范数约束的数据矩阵的奇异值分解模型,即

$$\begin{aligned}&\min_{\boldsymbol{U},\boldsymbol{V}}\ \|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_{\mathrm{F}}^2\\&\text{s.t.}\quad\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j=1\qquad j=1,2,\cdots,k\\&\qquad\|\boldsymbol{v}_j\|_q\leqslant t_j\qquad j=1,2,\cdots,k\end{aligned}\tag{3}$$

其中:$q=0$ 或 1;t_j 为载荷 $\boldsymbol{v}_j$ 的稀疏度参数。首先,该算法利用块坐标下降法将变量 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 按列分成 $2k$ 个块,即 $\boldsymbol{v}_1,\boldsymbol{u}_1,\boldsymbol{v}_2,\boldsymbol{u}_2,\cdots,\boldsymbol{v}_k,\boldsymbol{u}_k$,每个 $\boldsymbol{v}_j$(或 $\boldsymbol{u}_j,j=1,2,\cdots,k$)代表一个坐标块,并将目标函数重新表述为

$$\|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_{\mathrm{F}}^2=\Big\|\boldsymbol{X}-\sum_{i=1}^{k}\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}\Big\|_{\mathrm{F}}^2=\|\boldsymbol{E}_j-\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}\|_{\mathrm{F}}^2\qquad j=1,2,\cdots,k$$

http://bhxb.buaa.edu.cn

其中:$\boldsymbol{E}_j = \boldsymbol{X} - \sum_{i\neq j}\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}$;其次,当固定除 $\boldsymbol{v}_j$(和 $\boldsymbol{u}_j$,$j=1,2,\cdots,k$)以外的其他变量时,按照 $\boldsymbol{v}_1,\boldsymbol{u}_1,\boldsymbol{v}_2,\boldsymbol{u}_2,\cdots,\boldsymbol{v}_k,\boldsymbol{u}_k$ 的次序循环地求解关于 $\boldsymbol{v}_j$(和 $\boldsymbol{u}_j$)的子问题直至满足终止条件。BCD-SPCA 的优点是它可以求解具有特定稀疏度或非负的稀疏 PCA 问题并证明其是收敛的。缺点是当模型(3)中 $q=1$ 时,关于 $\boldsymbol{v}_j$ 的子问题的解是与参数 t_j 有关的软阈值函数,但求解其中的阈值的计算量较大,这使得子问题的求解变得复杂。

本文将模型(2)推广至秩 k 矩阵逼近,以便同时求 k 个稀疏主成分,此外将 $\boldsymbol{u}$ 是单位向量的约束还原成载荷 $\boldsymbol{v}_j$ 是单位向量的约束,并采用 BCD-SPCA 的分块模式,设计了求解 l_1 正则化秩 k 矩阵逼近问题的块坐标下降法。该算法的子问题的求解比 BCD-SPCA 简单。

2　秩 k 矩阵逼近的块坐标下降法

本节先提出了基于 l_1 正则化秩 k 矩阵逼近的稀疏 PCA 模型,并设计了求解该模型的块坐标下降法,同时还分析了该算法的计算复杂度与收敛性。

2.1　模型与算法

首先提出如下基于 l_1 正则化秩 k 矩阵逼近的稀疏 PCA 模型,即

$$\min_{\boldsymbol{U},\boldsymbol{V}} \ \|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \sum_{j=1}^{k}\gamma_j\|\boldsymbol{v}_j\|_1$$
$$\text{s.t.} \quad \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j = 1 \qquad j=1,2,\cdots,k \tag{4}$$

其中:正参数 $\gamma_j(j=1,2,\cdots,k)$ 的含义与模型(1)中的相同。由于求解既含有稀疏性的条件,又有正交性约束的模型是非常困难的[10],因此大多数的稀疏 PCA 模型并没有要求载荷的正交性,如模型(1)~模型(3)。鉴于此,在模型(3)中,取 $q=1$,把关于 $\boldsymbol{v}_j$ 的稀疏性条件放在目标中作为正则化项并保留 $\boldsymbol{v}_j$ 是单位向量的约束即得模型(4)。因此,模型(4)可看成是模型(3)的惩罚版本。同时,sPCA-rSVD-soft 的模型可以看作是模型(4)在 $k=1$ 时的特例(此时将约束条件变成 $\boldsymbol{u}$ 为单位向量即可)。

对上述模型,本文采用 BCD-SPCA 的分块模式,并利用块坐标下降法求解该问题。具体地,首先将目标函数等价表述为

$$\|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \sum_{j=1}^{k}\gamma_j\|\boldsymbol{v}_j\|_1 =$$
$$\|\boldsymbol{E}_j-\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}\|_{\mathrm{F}}^2 + \sum_{j=1}^{k}\gamma_j\|\boldsymbol{v}_j\|_1 \qquad j=1,2,\cdots,k$$

其中:$\boldsymbol{E}_j = \boldsymbol{X} - \sum_{i\neq j}\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}$。在原问题(4)中,当固定除 $\boldsymbol{v}_j$ 以外的其他坐标块的变量时,得到关于 $\boldsymbol{v}_j$ 的子问题,即

$$\min_{\boldsymbol{v}_j} \ \|\boldsymbol{E}_j-\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}\|_{\mathrm{F}}^2 + \gamma_j\|\boldsymbol{v}_j\|_1$$
$$\text{s.t.} \quad \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j = 1 \tag{5}$$

当固定除 $\boldsymbol{u}_j$ 以外的其他坐标块的变量时,得到关于 $\boldsymbol{u}_j$ 的子问题,即

$$\min_{\boldsymbol{u}_j}\|\boldsymbol{E}_j-\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}\|_{\mathrm{F}}^2 \tag{6}$$

为了方便表述,省去 $\boldsymbol{E}_j$、$\boldsymbol{u}_j$、$\boldsymbol{v}_j$、γ_j 的下角标,用 $\boldsymbol{E}$、$\boldsymbol{u}$、$\boldsymbol{v}$、γ 来表示。由于

$$\|\boldsymbol{E}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\|_{\mathrm{F}}^2 \ominus \|\boldsymbol{E}\|_{\mathrm{F}}^2 - 2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{v} + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}$$

则子问题(5)等价地表述为关于 $\boldsymbol{v}$ 在单位球上极小化二次函数加上 l_1 范数,即

$$\min_{\boldsymbol{v}\in\mathbf{R}^p} \ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} - 2(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})^{\mathrm{T}}\boldsymbol{v} + \gamma\|\boldsymbol{v}\|_1$$
$$\text{s.t.} \quad \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = 1 \tag{7}$$

因为 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}=1$,从而子问题(6)可等价表述为关于 $\boldsymbol{u}$ 的无约束优化问题,即

$$\min_{\boldsymbol{u}\in\mathbf{R}^n} \ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} - 2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{v} \tag{8}$$

定理 1 和定理 2 给出了子问题(7)和子问题(8)的显式解。

定理 1　当 $S_{\gamma e}(2\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})\neq 0$ 时,子问题(7)的解为

$$\boldsymbol{v}^* = S_{\gamma e}(2\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})/\|S_{\gamma e}(2\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})\|_2$$

证明　将问题(7)的约束条件代入问题(7)的目标函数,得到问题(7)的最优解与式(9)的相同。

$$\min_{\boldsymbol{v}\in\mathbf{R}^p} \ -2(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})^{\mathrm{T}}\boldsymbol{v} + \gamma\|\boldsymbol{v}\|_1$$
$$\text{s.t.} \quad \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = 1 \tag{9}$$

再根据 $S_{\gamma e}(2\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})\neq 0$ 和文献[7]中命题 14 类似的方法,可以证明式(9)有唯一的最优解

$$\boldsymbol{v} = S_{\gamma e}(2\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})/\|S_{\gamma e}(2\boldsymbol{E}^{\mathrm{T}}\boldsymbol{u})\|_2$$

证毕

定理 2　子问题(8)的最优解为 $\boldsymbol{u}^* = \boldsymbol{E}\boldsymbol{v}$。

BCD-sPCA-rSVD 就是利用上面所介绍的分块模式,并按照 $\boldsymbol{v}_1,\boldsymbol{u}_1,\boldsymbol{v}_2,\boldsymbol{u}_2,\cdots,\boldsymbol{v}_k,\boldsymbol{u}_k$ 的次序循环地求解关于 $\boldsymbol{v}_j$ 和 $\boldsymbol{u}_j$ 的子问题直至满足终止条件。BCD-sPCA-rSVD 的伪码描述见算法 1。

算法 1　稀疏 PCA 的 BCD-sPCA-rSVD 算法

输入:数据矩阵 $\boldsymbol{X}$,稀疏主成分的个数 k,稀疏度参数 $\gamma_j(j=1,2,\cdots,k)$。

1. 给定初始点 $\boldsymbol{V}^0,\boldsymbol{U}^0$
2. repeat
3. 　for $j=1,2,\cdots,k$ do
4. 　　计算 $\boldsymbol{E}_j^{\mathrm{T}}\boldsymbol{u}_j = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}_j - \sum_{i\neq j}\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j$

http://bhxb.buaa.edu.cn

5. 求解关于 $\boldsymbol{v}_j$ 的子问题(5)来更新 $\boldsymbol{v}_j$,即令 $\boldsymbol{v}_j = S_{\gamma e}(2\boldsymbol{E}_j^{\mathrm{T}}\boldsymbol{u}_j)/\|S_{\gamma e}(2\boldsymbol{E}_j^{\mathrm{T}}\boldsymbol{u}_j)\|_2$

6. 计算 $\boldsymbol{E}_j\boldsymbol{v}_j = \boldsymbol{X}\boldsymbol{v}_j - \sum_{i\neq j}\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j$

7. 求解关于 $\boldsymbol{u}_j$ 的子问题(6)来更新 $\boldsymbol{u}_j$,即令 $\boldsymbol{u}_j = \boldsymbol{E}_j\boldsymbol{v}_j$

8. end for

9. until stopping criterion satisfied

输出:稀疏主成分载荷 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k$。

由于 BCD-sPCA-rSVD 采用了与 BCD-SPCA 相同的分块模式,因此在算法 1 的第 9 行中,采用与 BCD-SPCA 相同的终止条件,即 $\boldsymbol{U}$ 或 $\boldsymbol{V}$ 的更新率小于给定阈值,或算法的迭代次数达到预设的最大迭代次数时,算法终止。

虽然 BCD-sPCA-rSVD 与第 1 节中所介绍的 3 种稀疏 PCA 算法都是块坐标下降法,但从模型和算法的角度考虑,它们又有一定的区别:

1)与 SPCA 相比,BCD-sPCA-rSVD 的分块模式与模型均与 SPCA 不同。

2)与 sPCA-rSVD 相比,BCD-sPCA-rSVD 能同时求多个主成分,并且模型的目标函数是含有 l_1 范数惩罚项的数据矩阵的秩 k 逼近,约束是要求 $\boldsymbol{v}_j$ 为单位向量而不是 $\boldsymbol{u}_j$。

3)与 BCD-SPCA 相比,BCD-sPCA-rSVD 的模型是把控制 PC 载荷的稀疏性条件作为目标函数的惩罚项,而不是作为约束条件。

2.2 计算复杂度分析

下面讨论 BCD-sPCA-rSVD 的计算复杂度。为了与 SPCA 进行比较,本文以算法 1 的第 3 行至第 8 行为一次迭代。

易知该算法的计算任务主要是计算第 4 行与第 6 行中的 $\boldsymbol{E}_j^{\mathrm{T}}\boldsymbol{u}_j$ 和 $\boldsymbol{E}_j\boldsymbol{v}_j$。设 J 为 $\boldsymbol{X}$ 中非零元素的个数。当计算 $\boldsymbol{E}_j^{\mathrm{T}}\boldsymbol{u}_j$ 时,需要计算矩阵向量乘积 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}_j$,其计算量为 J;此外还需要计算 $k-1$ 次向量内积 $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j$,其计算量为 $(k-1)n$。从而计算 $\boldsymbol{E}_j^{\mathrm{T}}\boldsymbol{u}_j$ 的总计算量是 $J+(k-1)n$。类似可得计算 $\boldsymbol{E}_j\boldsymbol{v}_j$ 的计算量是 $J+(k-1)p$。从而计算 $\boldsymbol{E}_j^{\mathrm{T}}\boldsymbol{u}_j$ 和 $\boldsymbol{E}_j\boldsymbol{v}_j$ 的总的计算复杂度是 $O(J+k\max(n,p))$。该算法每次迭代的计算复杂度为 $O(Jk+\max(n,p)\,k^2)$。

表 1 列出了各稀疏 PCA 算法每次迭代的计算复杂度(包括 BCD-sPCA-rSVD、SPCA 以及 BCD-SPCA)。这里还需强调的是,当 $n>p$ 时,SPCA 在第一次迭代的计算复杂度为 $np^2+O(p^3)$,之后每次迭代的计算复杂度均为 $O(p^3)$。由表 1 可看出:

1)不论对于 $n>p$ 还是 $n<p$ 的数据集,BCD-sPCA-rSVD 的计算复杂度关于样本个数 n 与变量维数 p 都是线性的。

表 1 各稀疏 PCA 算法的计算复杂度

Table 1 Computational complexity of each sparse PCA algorithm

算法	计算复杂度	
	$n>p$	$n<p$
BCD-sPCA-rSVD	$O(Jk+nk^2)$	$O(Jk+pk^2)$
SPCA	$O(p^3)$	$O(Jk+pk^2)$
BCD-SPCA	$O(Jk+nk^2)$	$O(Jk+pk^2)$

2)在高维低样本数据集($n<p$)下,虽然 3 个算法每次迭代的计算复杂度完全相同,但是 BCD-sPCA-rSVD 和 BCD-SPCA 是将变量分成 $2k$ 块逐次更新,而 SPCA 是分成两块逐次更新,理论上前面 2 个算法要比 SPCA 收敛得更快。后文 3.3 节和3.4 节的数值结果也很好地支持了该事实。进一步,BCD-sPCA-rSVD 在每次迭代又比 BCD-SPCA 少求解一个方程,所以 BCD-sPCA-rSVD 的计算效率更高。

2.3 收敛性分析

在分析 BCD-sPCA-rSVD 的收敛性时,首先利用约束的指示函数将原问题(4)重新表述为无约束优化问题,即令

$$f(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k,\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k) = \|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \sum_{j=1}^{k}\gamma_j\|\boldsymbol{v}_j\|_1 + \sum_{j=1}^{k}f_j(\boldsymbol{v}_j)$$

其中:$f_j(\boldsymbol{v}_j)(j=1,2,\cdots,k)$ 为指示函数,定义为

$$f_j(\boldsymbol{v}_j)=\begin{cases}0 & \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j = 1\\ \infty & \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j \neq 1\end{cases}$$

进一步,令

$$f_0(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k,\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k) = \|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_{\mathrm{F}}^2 = \left\|\boldsymbol{X}-\sum_{j=1}^{k}\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}\right\|_{\mathrm{F}}^2$$

$$g_0(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k) = \sum_{j=1}^{k}(\gamma_j\|\boldsymbol{v}_j\|_1 + f_j(\boldsymbol{v}_j))$$

则原问题(4)可等价表示为无约束优化问题

$$\min_{\{\boldsymbol{u}_j,\boldsymbol{v}_j\}_{j=1}^{k}} f_0(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k,\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k) + g_0(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k) \tag{10}$$

问题(10)的目标函数包括光滑的 $f_0(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k,\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k)$ 和非光滑的 $g_0(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k)$,其中非光滑部分关于 $\boldsymbol{v}_j$ 是可分离的。而本文提出的 BCD-sPCA-rSVD 可看作是求解问题(10)的块坐标下降法。令 $\boldsymbol{x}=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k,\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k)$,初始点 $\boldsymbol{x}^0=(\boldsymbol{u}_1^0,\boldsymbol{u}_2^0,\cdots,\boldsymbol{u}_k^0,\boldsymbol{v}_1^0,\boldsymbol{v}_2^0,\cdots,\boldsymbol{v}_k^0)$,其中 $\boldsymbol{v}_j^0(j=1,2,\cdots,k)$ 为主成分载荷,

http://bhxb.buaa.edu.cn

$\boldsymbol{u}_j^0 = \boldsymbol{X}\boldsymbol{v}_j^0$。定理 3 证明了 BCD-sPCA-rSVD 是收敛的。

定理 3　BCD-sPCA-rSVD 算法产生的点列能收敛到 f 的稳定点。

证明　①根据水平集的定义可知,$\boldsymbol{L}_0 = \{\boldsymbol{x}: f(\boldsymbol{x}) \leqslant f(\boldsymbol{x}^0), \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j = 1, j = 1,2,\cdots,k\}$,易知 $\boldsymbol{L}_0$ 是紧的且 $f(\boldsymbol{x})$ 在 $\boldsymbol{L}_0$ 上连续。②因为 f_0 的定义域为开集且在该定义域上 f_0 是 Gateaux 可微的,则根据文献[14]中的引理 3.1 中函数正则的判定条件,$f(\boldsymbol{x})$ 正则。③由定理 1 知子问题(7)有唯一解,因此子问题(5)也有唯一解;而子问题(6)的目标函数关于 $\boldsymbol{u}_j$ 是严格凸的,所以子问题(6)也有唯一解。根据条件①、②、③,并由文献[14]中的定理 4.1(c)可知,BCD-sPCA-rSVD 算法产生的点列收敛于 f 的稳定点。　证毕

3　数值结果与分析

为了验证 BCD-sPCA-rSVD 的可行性与有效性,本文将 BCD-sPCA-rSVD 应用到各种数据集中,包括真实数据集(Pitprop, Colon Cancer, 20Newsgroups)和合成数据,并与其他已有的稀疏 PCA 算法进行比较。本文中的所有数值实验均是在处理器为因特尔双核 2.20 GHz 的计算机上运行,实验所需的代码均利用 MATLAB 2010a 软件编写。

BCD-sPCA-rSVD 中,初始点 $\boldsymbol{V}^0$ 的每一列为主成分载荷,$\boldsymbol{U}^0 = \boldsymbol{X}\boldsymbol{V}$,控制最大迭代次数为 1 000,$\boldsymbol{U}$、$\boldsymbol{V}$的更新率均取 10^{-6}。如下各实验中,这些参数的取值均保持不变。

3.1　Pitprop 数据

本节利用 Jeffers 在 1967 年引进的 Pitprop 数据[15]来验证 BCD-sPCA-rSVD 算法的性能。Pitprop数据包含 180 个样本,13 个变量,即 $n = 180, p = 13$。它是主成分存在解释困难的最为经典的数据集,该数据集的前 6 个主成分载荷中没有零元素。本次实验与 BCD-sPCA-rSVD 比较的算法包括 SPCA[1]、SCoTLASS[2]、DSPCA[3]、sPCA-rSVD-soft[4]、Gpower($\mathrm{Gpower}_{l_1,k}$)[10]、ALSPCA[11]以及 BCD-SPCA[12]。对于 BCD-sPCA-rSVD,取原问题(4)中的 $\gamma_1 \sim \gamma_6$ 依次为 0.6, 0.6, 0.6, 1.0, 1.0, 1.0。

表 2 为 Pitprop 数据下 PCA 与各稀疏 PCA 算法的性能指标。其中第 2 列的稀疏度指载荷中零元素的个数(当元素绝对值小于阈值 0.001 时,即截断为 0)。第 3 列表示非正交性,设任意 2 个载荷 $\boldsymbol{v}_i$ 与 $\boldsymbol{v}_j$ 之间的夹角为 α_{ij},则非正交性为 $|90 - \alpha_{ij}|$的最大值。显然该值越小,表明正交性越好。第 4 列的相关性指各主成分之间相关系数绝对值的最大值。第 5 列的 PEV 指前 k 个稀疏主成分的调整方差和占所有主成分方差和的百分比,它是评价稀疏 PCA 算法的重要指标。本节采用 Zou 和 Hastie[1]提出的方法来计算稀疏主成分的 PEV。最后一列指数据矩阵 $\boldsymbol{X}$ 的相对重构误差 RRE,它是从数据恢复重构的角度来评价算法的性能。文献[12]给出了 RRE 的计算方法。

表 2　Pitprop 数据:各稀疏 PCA 算法的性能指标

Table 2　Pitprop data: Performance indicators of each sparse PCA algorithm

算法	稀疏度	非正交性	相关性	PEV/%	RRE/%
PCA	0	0	0	86.94	36.06
DSPCA	63	13.63	0.57	72.46	47.71
Gpower	63	17.88	0.51	75.04	45.89
SCoTLASS	27	0.32	0.44	78.24	49.24
ALSPCA	63	0	0.30	73.32	45.37
sPCA-rSVD-soft	53	14.76	0.46	76.59	46.76
SPCA	60	0.86	0.40	75.82	44.48
BCD-SPCA	63	20.05	0.40	75.86	44.19
BCD-sPCA-rSVD	63	1.51	0.28	75.13	44.18

表 2 表明,BCD-sPCA-rSVD 得到的主成分的载荷稀疏度高,有较小的相关系数与较高的 PEV,且非正交性的表现优于 DSPCA、Gpower、sPCA-rSVD-soft 与 BCD-SPCA;在数据恢复重构方面,BCD-sPCA-rSVD 的 RRE 值与 BCD-SPCA 和 SPCA 的近似相等,且明显小于其他稀疏 PCA 算法,进一步验证了同时求多个主成分可以降低计算所产生的累积误差这一事实。

3.2　合成数据

考虑 Zou 和 Hastie[1]提出的人工合成数据,本文首先要定义 3 个变量:$V_1 \sim N(0,290)$, $V_2 \sim N(0,300)$, $V_3 = -0.3V_1 + 0.925V_2 + \varepsilon$,其中 $\varepsilon \sim N(0,1)$、V_1、V_2 与 ε 独立。然后通过上述 3 个变量生成如下 10 个观测变量:

$$X_i = \begin{cases} V_1 + \varepsilon_i^1 & i = 1,2,3,4 \\ V_2 + \varepsilon_i^2 & i = 5,6,7,8 \\ V_3 + \varepsilon_i^3 & i = 9,10, \end{cases}$$

其中 $\varepsilon_i^j \sim N(0,1)$ 且 ε_i^j 是相互独立的($j = 1,2,3$; $i = 1,2,\cdots,10$)。由如上关系可计算出变量 V_1、V_2 与 V_3 的方差依次为 290、300 与 283.8。与这 3 个变量相关的观测变量的个数分别为 4、4、2,且最后 2 个观测变量仅与 V_3 有关,而 V_3 又与 V_1 和 V_2 线性相关,因此只需计算前 2 个主成分就能包含原始数据的大部分信息。同时作出如下推测:

http://bhxb.buaa.edu.cn

1）通过观测变量 X_5、X_6、X_7、X_8 来计算对应于 V_2 的第一稀疏主成分 PC1。

2）通过观测变量 X_1、X_2、X_3、X_4 来计算对应于 V_1 的第二稀疏主成分 PC2。

3）由于(X_1,X_2,X_3,X_4)与(X_5,X_6,X_7,X_8)是相互独立的,则它们所对应的稀疏主成分不相关,并且其载荷是相互正交的。

与 Zou 和 Hastie[1] 的处理方式相同,本文利用精确的协方差矩阵进行计算。此外,为了使算法能够更好地去掉主成分之间的相关性,同 ALSPCA[11] 的处理方式,先对数据进行预处理,即将协方差矩阵除以 $\max\limits_{i=1,2,\cdots,10} D(X_i)$,其中 $D(X_i)$ 表示 X_i 的方差。在模型(4)中取 γ_1 与 γ_2 均为 1。由于 SPCA、BCD-SPCA 以及 BCD-sPCA-rSVD 都是同时求解多个主成分的块坐标下降法,因此本实验及后面的 2 个实验均将 BCD-sPCA-rSVD 的计算结果与这 2 种算法的计算结果进行比较。表3为合成数据下 PCA 与各稀疏 PCA 算法的载荷。

表 3 合成数据:PCA 与各稀疏 PCA 算法的载荷

Table 3 Synthetic data: Loadings of PCA and each sparse PCA algorithm

变量	PCA		BCD-sPCA-rSVD		SPCA(BCD-SPCA)	
	PC1	PC2	PC1	PC2	PC1	PC2
X_1	-0.1163	-0.4784	0	-0.5000	0	-0.5000
X_2	-0.1162	-0.4784	0	-0.5000	0	-0.5000
X_3	-0.1162	-0.4784	0	-0.5000	0	-0.5000
X_4	-0.1162	-0.4783	0	-0.5000	0	-0.5000
X_5	0.3951	-0.1453	0.5000	0	0.5000	0
X_6	0.3951	-0.1453	0.5000	0	0.5000	0
X_7	0.3951	-0.1454	0.5000	0	0.5000	0
X_8	0.3951	-0.1453	0.5000	0	0.5000	0
X_9	0.4009	0.0091	0	0	0	0
X_{10}	0.4009	0.0091	0	0	0	0
PEV/%	99.72		80.46		80.46	

通过表 3 发现,BCD-sPCA-rSVD 同其他2 种算法一样,成功地验证了上述全部 3 个预测,进一步验证了 BCD-sPCA-rSVD 是可行且有效的。

3.3 Colon Cancer 数据

结肠癌(Colon Cancer)基因表达数据集[16] 是典型的高维低样本数据。它包含 62 个样本(22 个正常组织样本和 40 个癌变组织样本)和2 000个基因表达变量,即 $n=62$,$p=2\,000$。

由于前 3 个主成分的 PEV 已接近 60%,因此在本次实验中取 $k=3$。对于这 3 种稀疏 PCA 算法,参考 ALSPCA[11] 的处理方式,先对数据进行预处理,令 $\boldsymbol{X}=\dfrac{\boldsymbol{X}}{\sqrt{n-1}}$。对于 BCD-sPCA-rSVD,取模型(4)中的 γ_1、γ_2、γ_3 分别为 1 350、1 700、1 700。为了验证 BCD-sPCA-rSVD 在高维数据集下的有效性,为本实验及下一个实验增加了 2 个体现效率的指标:计算时间和迭代次数,表 4 为 Colon Cancer 数据下各稀疏 PCA 算法的性能指标。

表 4 Colon Cancer 数据:各稀疏 PCA 算法的性能和效率指标

Table 4 Colon Cancer data: Performance and efficiency indicators of each sparse PCA algorithm

算法	稀疏度	非正交性	相关性	PEV/%	RRE/%	计算时间/s	迭代次数
PCA	0	0	0	58.35	64.54	1.98	
SPCA	5 370	22.88	0.49	46.12	65.48	4.47	165
BCD-SPCA	5 373	29.15	0.54	48.88	69.73	3.65	97
BCD-sPCA-rSVD	5 376	24.02	0.48	47.24	65.34	2.76	91

由表 4 可知,BCD-rSVD 所得的载荷在保持较高稀疏度与 PEV 情况下,需要较少的计算时间和迭代次数,并在非正交性、相关性、RRE 方面与其余 2 种算法的表现接近一致。这些事实很好地支持了表 1 中的复杂度分析结论,验证了该算法在解决高维稀疏 PCA 问题时的简单有效。

3.4 20Newsgroups 数据

20Newsgroups 是一组低维高样本的数据集,它记录了 100 个单词在16 242 篇报导中出现的频率,即 $n=16\,242$,$p=100$。所有的新闻报导均从全球最大的电子布告栏系统 Usenet 上取得。该数据集下载自 http://cs.nyu.edu/roweis/data.html。

在本次实验中取 $k=2$。与上一个实验相同,先对数据进行预处理,令 $\boldsymbol{X}=\dfrac{\boldsymbol{X}}{\sqrt{n-1}}$。对于 BCD-sPCA-rSVD,取模型(4)中的 γ_1、γ_2 分别为 166、166。表 5 为 20Newsgroups 数据下各稀疏 PCA 算法的性能指标。

表 5 20Newsgroups 数据:各稀疏 PCA 算法的性能和效率指标

Table 5 20Newsgroups data: Performance and efficiency indicators of each sparse PCA algorithm

算法	稀疏度	非正交性	相关性	PEV/%	RRE/%	时间/s	迭代次数
PCA	0	0	0	10.69	94.50	1.08	
SPCA	161	0.07	0.15	8.42	95.76	2.62	192
BCD-SPCA	161	17.64	0.30	8.71	95.34	3.11	58
BCD-sPCA-rSVD	166	0.02	0.09	8.58	95.60	2.73	52

由表 5 可知,BCD-sPCA-rSVD 获得的载荷在稀疏度、PEV、RRE 方面与其余 2 种算法的表现接

http://bhxb.buaa.edu.cn

近一致,但在非正交性与相关性方面的表现要优于其余 2 种算法,并且需要的计算时间和迭代次数最少,进一步验证了该算法在解决大规模稀疏 PCA 问题时的有效性。

4　结　论

本文首次提出了基于正则化秩 k 矩阵逼近的稀疏 PCA 模型(4),并设计了求解该模型的块坐标下降法(BCD-sPCA-rSVD)。BCD-sPCA-rSVD 不仅易于实现,理论分析与数值模拟还表明:

1)该算法减少了计算的累积误差。

2)该算法是收敛的。

3)该算法每次迭代的计算复杂度关于样本个数 n 与变量维数 p 都是线性的,因此它可以有效地求解大规模稀疏 PCA 问题。

参考文献(References)

[1] ZOU H, HASTIE T. Sparse principal component analysis[J]. Journal of Computational and Graphical Statistics, 2006, 15(2):265-286.

[2] TRENDAFILOV N T, JOLLIFFE I T. Projected gradient approach to the numerical solution of the SCoTLASS[J]. Computational Statistics and Data Analysis, 2006, 50(1):242-253.

[3] D'ASPREMONT A, GHAOUI L E, JORDAN M I, et al. A direct formulation for sparse PCA using semidefinite programming [J]. SIAM Review, 2007, 48(3):434-448.

[4] SHEN H, HUANG J Z. Sparse principal component analysis via regularized low rank matrix approximation[J]. Journal of Multivariate Analysis, 2008, 99(6):1015-1034.

[5] MOGHADDAM B, WEISS Y, AVIDAN S. Spectral bounds for sparse PCA: Exact and greedy algorithms[C]// Advances in Neural Information Processing Systems. Montreal: Neural Information Processing System Foundation, 2006:915-922.

[6] D'ASPREMONT A, BACH F R, GHAOUI L E. Optimal solutions for sparse principal component analysis[J]. Machine Learning, 2008, 9(7):1269-1294.

[7] LUSS R, TEBOULLE M. Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint[J]. SIAM Review, 2013, 55(1):65-98.

[8] LUSS R, TEBOULLE M. Convex approximations to sparse PCA via Lagrangian duality[J]. Operations Research Letters, 2011, 39(1):57-61.

[9] SIGG C, BUHMANN J. Expectation-maximization for sparse and nonnegative PCA[C]// Proceedings of the 25th International Conference on Machine Learning. NewYork: ACM, 2008: 960-967.

[10] JOURNEE M, NESTEROV Y, RICHTARIK P, et al. Generalized power method for sparse principal component analysis [J]. Journal of Machine Learning Research, 2010, 11(2): 517-553.

[11] LU Z S, ZHANG Y. An augmented Lagrangian approach for sparse principal component analysis[J]. Math Program Series A, 2012, 135(1-2):149-193.

[12] ZHAO Q, MENG D Y, XU Z B, et al. A block coordinates descent approach for sparse principal component analysis[J]. Neurocomputing, 2015, 153(4):180-190.

[13] TIBSHIRANI R. Regression shrinkage and selection via the LASSO[J]. Journal of the Royal Statistical Society Series B, 1996, 58(3):267-268.

[14] TSENG P. Convergence of a block coordinates descent method for nondifferentiable minimization[J]. Journal of Optimization Theory Apply, 2001, 109(3):475-494.

[15] JEFFERS J N R. Two case studies in the application of principal component analysis[J]. Applied Statistics, 1967, 16(3): 225-236.

[16] ALON U, BARKAI N, NOTTERMAN D A, et al. Broad patterns of gene expression revealed by clustering of tumor and normal colon tissues probed by oligonucleotide arrays[J]. Proceedings of the National Academy of Sciences of the United States of America, 1999, 96(12):6745-6750.

作者简介:

杨茜　女,硕士研究生。主要研究方向:数值最优化及其应用。

刘红英　女,博士,副教授,硕士生导师。主要研究方向:数值最优化及其应用。

http://bhxb.buaa.edu.cn

Sparse principal component analysis via regularized rank-k matrix approximation

YANG Qian, LIU Hongying*

(School of Mathematics and Systems Science, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: In calculating the sparse principal components (PCs), attaining k PCs simultaneously can reduce the accumulated error arising from the calculation process. We proposed the sparse principal component model via regularized rank-k matrix approximation and designed a block coordinate descent method (BCD-sPCA-rSVD) to solve this problem. Its main idea is to first divide variables into $2k$ blocks by coordinates, and then solve sub-problem with respect to each single coordinate block when keeping other $2k-1$ variables fixed. By solving these sub-problems with explicit solutions recursively until the stopping criterion is satisfied, the BCD-sPCA-rSVD algorithm can be easily constructed. Its per-iteration complexity is linear in both sample size and variable dimensionality. The algorithm is convergent and easy to implement. Numerical simulation results show that the algorithm is feasible and effective when applied to real and synthetic data sets. The proposed method reduces the accumulated error and has lower computational complexity, which makes it well suited to handling large-scale problems of sparse principal component analysis.

Keywords: dimension reduction; sparse principal component; regularization; block coordinate descent method; singular value decomposition; threshold

http://bhxb.buaa.edu.cn

Received: 2016-05-31; **Accepted**: 2016-09-02; **Published online**: 2016-10-10 09:03

URL: www.cnki.net/kcms/detail/11.2625.V.20161010.0903.004.html

Foundation items: National Natural Science Foundation of China (61172060, 61403011)

* **Corresponding author**. E-mail: liuhongying@buaa.edu.cn