关于 SVD 与 PCA 等价性的研究

吴春国[1)]　梁艳春[1),2)]　孙延风[1)]　周春光[1)]　吕英华[3)]

[1)](吉林大学计算机科学与技术学院，教育部符号计算与知识工程重点实验室　长春 130012)

[2)](中国科学院计算技术研究所智能信息处理重点实验室　北京 100080)

[3)](东北师范大学计算机科学系　长春 130024)

摘　要　利用矩阵的 Frobenius 范数对奇异值分解(Singular Value Decomposition, SVD)的正规正交基的最优性给出了一种新的证明,论述了 SVD 与主成分分析的等价性.

关键词　正常正交分解;主成分分析;Karhunen-Loève 分解;奇异值分解

中图法分类号　TP183

On the Equivalence of SVD and PCA

WU Chun-Guo[1)]　LIANG Yan-Chun[1),2)]　SUN Yan-Feng[1)]　ZHOU Chun-Guang[1)]　LU Ying-Hua[3)]

[1)](*College of Computer Science and Technology*, *Key Laboratory for Symbolic Computation and Knowledge Engineering of Ministry of Education*, *Jilin University*, *Changchun* 130012)

[2)](*Key Laboratory of Intelligent Information Processing*, *Institute of Computing Technology*, *Chinese Academy of Sciences*, *Beijing* 100080)

[3)](*Department of Computer Science*, *Northeast Normal University*, *Changchun* 130024)

Abstract　A novel proof on the optimal performance of the proper orthogonal bases of the singular value decomposition (SVD) is presented based on the Frobenius norm. The equivalence of the SVD and the principal component analysis (PCA) is indicated.

Keywords　proper orthogonal decomposition; principal component analysis; Karhunen-Loève decomposition; singular value decomposition

1　引　言

在图像处理、自适应控制、信号分析等很多领域都会涉及到数据压缩和特征抽取的问题,这类问题的解决多是依赖于正常正交分解 (Proper Orthogonal Decomposition, POD) 的思想[1,2].实现正常正交分解的方法有多种,通常较为常用的有主成分分析 (Principal Component Analysis, PCA)、K-L 分解 (Karhunen-Loève Decomposition, KLD)和奇异值分解(Singular Value Decomposition, SVD)[3~5].有关这三种方法在不同领域的应用研究日渐增多,例如,文献[6]介绍了 POD 方法应用于微机电系统模型缩减的近期研究进展,给出了基于神经网络实现 PCA 的方法以及在微机电系统动力建模方面的应用举例,并与 KLD 方法的有关结果进行了比较研究,验证了这两种 POD 方法在特征抽取效果上的等价性,分析了在算法实现上的若干差异.文献[7]提出了基于 SVD 和数据融合的脸像鉴别方法,首先利用 SVD 方法,求出脸像矩阵的奇异值和奇异值向量,以它们作为特征矢量进行脸像鉴别,分别得出了基于奇异值和重建误差的更为准确的脸像鉴别结果.文献[8]将 SVD 引入到模糊模型结构分析问题中,研究结果表明,将 SVD 与模糊模型统计信息准则相结合来指导模糊规则的删除与合并,可从模型结构的精简性、模型拟合以及泛化性能等方面综合

收稿日期:2002-08-22;修改稿收到日期:2003-06-12.本课题得到吉林省科技发展计划项目(20030520)、教育部科学技术研究重点项目(02090)和国家自然科学基金(60175024)资助.**吴春国**,男,1976 年生,博士研究生,主要从事计算机应用和数值计算方法等方面的研究.**梁艳春**,男,1953 年生,博士,教授,博士生导师,主要从事人工神经网络和进化计算等方面的研究.**孙延风**,男,1972 年生,博士研究生,主要从事人工神经网络和数值计算方法等方面的研究.**周春光**,男,1947 年生,博士,教授,博士生导师,主要从事人工神经网络和进化计算等方面的研究.**吕英华**,男,1962 年生,教授,主要从事计算机应用研究.

地确定最优模型结构.与这三种 POD 方法的应用研究相比,对于它们的理论研究却为数不多,尤其缺少有关这三种方法等价性的系统完整的论述.文献[9]对离散情况下 KLD 与 PCA 的等价性给出了证明.文献[1]按照求解优化问题极值的方式证明了 SVD 与 PCA 的等价性.本文将在文献[1]的基础上借助矩阵的 Frobenius 范数从新的角度对 SVD 与 PCA 的等价性给出证明.

2　SVD 与 PCA 的等价性

设有 R^m 空间的样本集 $\{\boldsymbol{x}_i\}_{i=1}^n$,样本矩阵为 $\boldsymbol{X}$($\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]$).文献[1]已经从矩阵特征值问题和样本渐进性关系方面论证了 SVD 与 PCA 的等价性,利用 Lagrange 函数证明了 SVD 的正规正交基满足 POD 的最优性要求.下面借助矩阵的 Frobenius 范数对 SVD 的正规正交基的最优性给出一种新的证明,从而证明 SVD 与 PCA 的等价性.在以下的定理中,$\|\cdot\|_2$ 表示向量的欧氏范数,$\|\cdot\|_F$ 表示矩阵的 Frobenius 范数.

引理 1[10].　设秩为 r 的 $n\times m(m<n)$ 阶矩阵 $\boldsymbol{A}$ 的奇异值分解是 $\boldsymbol{U\Sigma V}^{\mathrm{T}}$,其中 $\boldsymbol{U},\boldsymbol{V}$ 是正交阵,且 $\boldsymbol{U}\in R^{n\times n}$,$\boldsymbol{V}\in R^{m\times m}$

$$\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_m \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0\end{pmatrix} \tag{1}$$

奇异值按降序排列为 $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>\sigma_{r+1}=\cdots=\sigma_m=0$,则对任意的 $l\leq r$,有

$$\min_{\mathrm{rank}(\boldsymbol{B})=l}\|\boldsymbol{A}-\boldsymbol{B}\|_F^2=\|\boldsymbol{A}-\boldsymbol{A}_l\|_F^2=\sum_{i=l+1}^{r}\sigma_i^2 \tag{2}$$

其中

$$\boldsymbol{A}_l=\sum_{i=1}^{l}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}} \tag{3}$$

下面我们给出 SVD 的正规正交基满足 POD 最优性的定理与证明.

定理 1.　设 $\boldsymbol{X}^{\mathrm{T}}$ 为样本矩阵的转置,$\boldsymbol{V}$ 是对 $\boldsymbol{X}^{\mathrm{T}}$ 进行奇异值分解获得的样本空间的正规正交基,$\boldsymbol{\Phi}$ 是样本空间任意一组正规正交基,分别取 $\boldsymbol{V},\boldsymbol{\Phi}$ 的前 l 个基向量对样本矩阵进行重构,前者的重构误差记为 $\varepsilon^2(\boldsymbol{V}_l)$,后者的重构误差记为 $\varepsilon^2(\boldsymbol{\Phi}_l)$,则

$$\varepsilon^2(\boldsymbol{V}_l)\leq\varepsilon^2(\boldsymbol{\Phi}_l) \tag{4}$$

证明.　用 $\hat{\boldsymbol{x}}_j,\tilde{\boldsymbol{x}}_j(j=1,2,\cdots,n)$ 分别表示以 $\boldsymbol{V}$,$\boldsymbol{\Phi}$ 的前 l 个基向量对原来第 j 个样本的重构,即

$$\hat{\boldsymbol{x}}_j=\sum_{i=1}^{l}\alpha_{ij}\boldsymbol{v}_i,\ \tilde{\boldsymbol{x}}_j=\sum_{i=1}^{l}\beta_{ij}\boldsymbol{\varphi}_i,\ j=1,2,\cdots,n \tag{5}$$

其中

$$\alpha_{ij}=\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{x}_j,\ \beta_{ij}=\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{x}_j \tag{6}$$

重构误差分别记为

$$\varepsilon^2(\boldsymbol{V}_l)=\sum_{j=1}^{n}\|\boldsymbol{x}_j-\hat{\boldsymbol{x}}_j\|_2^2=\|\boldsymbol{X}-\hat{\boldsymbol{X}}\|_F^2 \tag{7}$$

$$\varepsilon^2(\boldsymbol{\Phi}_l)=\sum_{j=1}^{n}\|\boldsymbol{x}_j-\tilde{\boldsymbol{x}}_j\|_2^2=\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_F^2 \tag{8}$$

其中 $\hat{\boldsymbol{X}}=[\hat{\boldsymbol{x}}_1,\hat{\boldsymbol{x}}_2,\cdots,\hat{\boldsymbol{x}}_n]$,$\tilde{\boldsymbol{X}}=[\tilde{\boldsymbol{x}}_1,\tilde{\boldsymbol{x}}_2,\cdots,\tilde{\boldsymbol{x}}_n]$.

由引理知,对于任意的矩阵 $\boldsymbol{C}$,若它满足 $\boldsymbol{C}\in R^{n\times m}$,且 $\mathrm{rank}(\boldsymbol{C})=s\leq l$,则有

$$\|\boldsymbol{A}-\boldsymbol{C}\|_F^2\geq\min_{\mathrm{rank}(\boldsymbol{B})=s}\|\boldsymbol{A}-\boldsymbol{B}\|_F^2=\|\boldsymbol{A}-\boldsymbol{A}_s\|_F^2$$

$$=\sum_{i=s+1}^{r}\sigma_i^2\geq\sum_{i=l+1}^{r}\sigma_i^2=\|\boldsymbol{A}-\boldsymbol{A}_l\|_F^2 \tag{9}$$

事实上,因为 $\mathrm{rank}(\boldsymbol{C})=s$,并且 $\min\limits_{\mathrm{rank}(\boldsymbol{B})=s}\|\boldsymbol{A}-\boldsymbol{B}\|_F^2$ 是所有秩数等于 s 的矩阵与 $\boldsymbol{A}$ 的差的最小 Frobenius 范数,所以有 $\|\boldsymbol{A}-\boldsymbol{C}\|_F^2\geq\min\limits_{\mathrm{rank}(\boldsymbol{B})=s}\|\boldsymbol{A}-\boldsymbol{B}\|_F^2$ 成立.此时,注意到

$$\varepsilon^2(\boldsymbol{V}_l)=\|\boldsymbol{X}-\hat{\boldsymbol{X}}\|_F^2=\|(\boldsymbol{X}-\hat{\boldsymbol{X}})^{\mathrm{T}}\|_F^2=\|\boldsymbol{X}^{\mathrm{T}}-\hat{\boldsymbol{X}}^{\mathrm{T}}\|_F^2 \tag{10}$$

$$\varepsilon^2(\boldsymbol{\Phi}_l)=\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_F^2=\|(\boldsymbol{X}-\tilde{\boldsymbol{X}})^{\mathrm{T}}\|_F^2=\|\boldsymbol{X}^{\mathrm{T}}-\tilde{\boldsymbol{X}}^{\mathrm{T}}\|_F^2 \tag{11}$$

若以 $\boldsymbol{X}^{\mathrm{T}},\tilde{\boldsymbol{X}}^{\mathrm{T}},\hat{\boldsymbol{X}}^{\mathrm{T}}$ 分别替换式(9)中的 $\boldsymbol{A},\boldsymbol{C},\boldsymbol{A}_l$,则欲证明

$$\varepsilon^2(\boldsymbol{V}_l)\leq\varepsilon^2(\boldsymbol{\Phi}_l),$$

只需证明

$$\mathrm{rank}(\tilde{\boldsymbol{X}}^{\mathrm{T}})\leq l \tag{12}$$

并且

$$\hat{\boldsymbol{X}}^{\mathrm{T}}=\sum_{i=1}^{l}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}} \tag{13}$$

由 $\tilde{x}_j\in\mathrm{span}(\varphi_1,\varphi_2,\cdots,\varphi_l)$ $(j=1,2,\cdots,n)$,式(12)显然成立 .

下面证明式(13).假设对样本矩阵的转置 $\boldsymbol{X}^{\mathrm{T}}$ 有如引理中形式的奇异值分解,即有 $\boldsymbol{X}^{\mathrm{T}}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$,$\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_m)$,且 $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>\sigma_{r+1}=\cdots=\sigma_m=0$. 易知 $\boldsymbol{X}^{\mathrm{T}}$ 可以写成

$$\boldsymbol{X}^{\mathrm{T}}=\sum_{i=1}^{r}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}} \tag{14}$$

上式右端写成矩阵的形式有

$$\boldsymbol{X}^{\mathrm{T}}=\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^{\mathrm{T}} \tag{15}$$

其中 $\boldsymbol{U}_r=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r]$,$\boldsymbol{\Sigma}_r=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$,$\boldsymbol{V}_r^{\mathrm{T}}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r]^{\mathrm{T}}$.由式(15)可知

$$\boldsymbol{U}_r=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1} \tag{16}$$

即

$$\boldsymbol{U}_r=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]^{\mathrm{T}}[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r]\cdot$$

$$\begin{pmatrix}\sigma_1^{-1} & & & \\ & \sigma_2^{-1} & & \\ & & \ddots & \\ & & & \sigma_r^{-1}\end{pmatrix} \tag{17}$$

将上式两端右乘 $\boldsymbol{\Sigma}_r$,得

$$[\boldsymbol{u}_1\sigma_1, \boldsymbol{u}_2\sigma_2, \cdots, \boldsymbol{u}_r\sigma_r] = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{bmatrix} [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r] \tag{18}$$

由式(18)与式(6)可知 $\sigma_i u_{ji} = \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{v}_i = \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{x}_j = \alpha_{ij}$,再由式(5),有

$$\begin{aligned} &[\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \cdots, \hat{\boldsymbol{x}}_n] \\ &= \left[\sum_{i=1}^{l} \sigma_i u_{1i} \boldsymbol{v}_i, \sum_{i=1}^{l} \sigma_i u_{2i} \boldsymbol{v}_i, \cdots, \sum_{i=1}^{l} \sigma_i u_{ni} \boldsymbol{v}_i\right] \\ &= \sum_{i=1}^{l} \sigma_i \boldsymbol{v}_i \boldsymbol{u}_i^{\mathrm{T}}, \end{aligned}$$

对上式两边同时取转置,即得式(13).

综上证明,有

$$\begin{aligned} \|\boldsymbol{X}^{\mathrm{T}} - \widetilde{\boldsymbol{X}}^{\mathrm{T}}\|_F^2 &\geq \min_{\operatorname{rank}(\boldsymbol{B})=l} \|\boldsymbol{X}^{\mathrm{T}} - \boldsymbol{B}\|_F^2 \\ &= \|\boldsymbol{X}^{\mathrm{T}} - \hat{\boldsymbol{X}}^{\mathrm{T}}\|_F^2 = \sum_{i=l+1}^{r} \sigma_i^2 \end{aligned} \tag{19}$$

即

$$\varepsilon^2(\boldsymbol{V}_l) \leq \varepsilon^2(\boldsymbol{\Phi}_l)$$ 证毕.

至此,我们采用与文献[1]中不同的方法,完成了 SVD 正规正交基满足 POD 最优性的证明.结合文献[1]中的结论可知:SVD 方法中用以求解特征值问题的矩阵是 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$,PCA 中求解特征值问题的相关矩阵是 $\boldsymbol{R}_x$.通常 $\boldsymbol{R}_x$ 无法确定,是以样本矩阵来构造它的近似 $\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$,当样本数 n 增大时,有 $\boldsymbol{R}_x = \lim_{n\to\infty}\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$.而 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 与 $\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 具有相同的特征值和特征向量,因此 SVD 与 PCA 具有渐进关系.由于奇异值的平方就是原特征值问题中求得的特征值,所以从上面 SVD 最优性的证明中可以看出,SVD 与 PCA 都取前 l 个正规正交基对原来的样本做近似表达时,其误差也是相同的.因此,SVD 与 PCA 具有等价性.

3 结论与讨论

本文利用矩阵的 Frobenius 范数,以一种新的方式给出了 SVD 正规正交基满足 POD 最优性的证明,在文献[1]的基础上论述了 SVD 与 PCA 的等价性.利用不同途径、采用不同方式研究各种 POD 方法之间的等价性问题,对于完善 POD 的理论研究,推动 POD 的应用研究,是十分有益的.本文工作可以作为对于已有研究结果的补充.

参 考 文 献

1 Liang Y. C., Lee H. P., Lim S. P., Lin W. Z., Lee K. H., Wu C. G.. Proper orthogonal decomposition and its application —— Part I: Theory. Journal of Sound and Vibration, 2002, 252(3): 527～544

2 Liang Y. C., Lin W. Z., Lee H. P., Lim S. P., Lee K. H., Sun H.. Proper orthogonal decomposition and its application —— Part II: Model reduction for MEMS dynamical analysis. Journal of Sound and Vibration, 2002, 256(3): 515～532

3 Chatterjee A.. An introduction to the proper orthogonal decomposition. Current Science, 2000, 78(7): 808～817

4 Kunisch K., Volkwein S.. Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition. Journal of Optimization Theory and Applications, 1999, 102(2): 345～371

5 Holmes P., Lumley J. L., Berkooz G.. Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge: Cambridge University Press, 1996

6 Liang Y. C., Lin W. Z., Lee H. P., Lim S. P., Lee K. H., Feng D. P.. A neural-network-based method of model reduction for dynamic simulation of MEMS. Journal of Micromechanics and Microengineering, 2001, 11(3): 226～233

7 Wang Y. H., Tan T. N., Zhu Y.. Face identification based on singular value decomposition and data fusion. Chinese Journal of Computers, 2000, 23(6): 649～653(in Chinese)
(王蕴红,谭铁牛,朱 勇. 基于奇异值分解和数据融合的脸像鉴别. 计算机学报, 2000, 23(6): 649～653)

8 Yu J. S., Liu S. R.. Model construction optimization for a class of fuzzy models. Chinese Journal of Computers, 2001, 24(2): 164～172(in Chinese)
(俞金寿, 刘士荣. 一类模糊模型的结构优化问题研究. 计算机学报, 2001, 24(2): 164～172)

9 Bian Z. Q.. Pattern Recognition. Beijing: Tsinghua University Press, 1988(in Chinese)
(边肇祺. 模式识别. 北京:清华大学出版社, 1988)

10 Diamantaras K. I., Kung S. Y.. Principal Component Neural Networks Theory and Applications. New York: John Wiley & Sons Inc, 1996

WU Chun-Guo, born in 1976, Ph. D. candidate. His research interests include computer application and numerical computation.

LIANG Yan-Chun, born in 1953, Ph. D., professor and Ph. D. supervisor. His research interests include artificial neural network and evolutionary computation.

SUN Yan-Feng, born in 1972, Ph. D. candidate. His research interests include artificial neural network and numerical computation.

ZHOU Chun-Guang, born in 1947, Ph. D., professor and Ph. D. supervisor. His research interests include artificial neural network and evolutionary computation.

LU Ying-Hua, born in 1962, professor. His research interests focus on computer application.