一种新的联合对角化算法 [1]

岳　博　　焦李成

(西安电子科技大学雷达信号处理重点实验室　西安　710071)

摘　要　该文提出了一种用于对一组实对称正定矩阵进行联合对角化的新算法. 文中的研究表明, 对一组实对称正定矩阵的联合对角化问题可以近似成一个独立分量分析问题来解决, 并且给出了相应的算法. 同时, 算法的性能也通过一个实例与其它方法进行了比较.

关键词　联合对角化, 独立分量分析, 高斯分布

中图号　TN911.72

1 引　言

假设 K 个 $N\times N$ 维的实对称正定矩阵 $\boldsymbol{R}_1,\cdots,\boldsymbol{R}_K$ 具有相同的特征向量, 即存在正交矩阵 $\boldsymbol{U}$ 和对角矩阵 $\boldsymbol{\Lambda}_1,\cdots,\boldsymbol{\Lambda}_K$, 使得 $\forall k,\ k=1,\cdots,K,\ \boldsymbol{R}_k=\boldsymbol{U}\boldsymbol{\Lambda}_k\boldsymbol{U}^T$. 当给定 $\boldsymbol{R}_1,\cdots,\boldsymbol{R}_K$ 时, 我们需要估计矩阵 $\boldsymbol{U}$. 文献 [1] 指出, 这一矩阵的联合对角化问题当 $K>2$ 时只能得到近似解, 并且能够通过求解下面的优化问题:

$$\arg\min_{\boldsymbol{A}}\sum_{k=1}^{K}h_k[\log\det(\mathrm{diag}(\boldsymbol{A}\boldsymbol{R}_k\boldsymbol{A}^T))-\log\det(\boldsymbol{A}\boldsymbol{R}_k\boldsymbol{A}^T)] \tag{1}$$

得到 $\boldsymbol{U}=\boldsymbol{A}^{-1}$. 上式中的 $h_k,k=1,\cdots,K$ 是对应于 $\boldsymbol{R}_k$ 的权值, 我们假设它是归一化的, 即 $\sum_{k=1}^{K}h_k=1$, $\mathrm{diag}(\boldsymbol{A}\boldsymbol{R}_k\boldsymbol{A}^T)$ 表示由矩阵 $\boldsymbol{A}\boldsymbol{R}_k\boldsymbol{A}^T$ 的对角线元素组成的对角矩阵, $\det(\cdot)$ 表示取矩阵的行列式, T 表示向量或矩阵的转置.

通过首先对 $\boldsymbol{R}_1,\cdots,\boldsymbol{R}_K$ 的均值 $\sum_{k=1}^{K}h_k\boldsymbol{R}_k$ 的对角化, 我们能够使 (1) 式中的 $\boldsymbol{A}$ 成为一个正交矩阵 [2] . 这样在 (1) 式求和中的第二项将是一个常数, 因此我们只需选择正交矩阵 $\boldsymbol{A}$ 使得包含第一项的和最小化.

对于任意的实对称正定矩阵 $\boldsymbol{R}$, 我们可以引入一个随机变量 $\boldsymbol{z}$, 使得 $\boldsymbol{z}\sim N(\boldsymbol{R})$, $N(\boldsymbol{R})$ 表示均值为 0 , 协方差矩阵为 $\boldsymbol{R}$ 的高斯分布. 因此在本文中, 我们首先引入一个值域为 $\{1,2,\cdots,K\}$ 的离散随机变量 q , 并且有 $P(q=k)=h_k$. 一个 $\Re^N$ 上的随机向量 $\boldsymbol{x}$, 并且有 $p(\boldsymbol{x}|q=k)=N(\boldsymbol{R}_k)$, 这样 $\boldsymbol{x}$ 的概率密度为

$$p(\boldsymbol{x})=\sum_{k=1}^{K}h_kN(\boldsymbol{R}_k) \tag{2}$$

它是一个混合高斯模型的形式. 由此, 我们有条件熵 $H(\boldsymbol{x}|q)=\frac{1}{2}\sum_{k=1}^{K}h_k\log\det(2\pi\boldsymbol{R}_k)+\frac{N}{2}$. 对于 $\boldsymbol{x}$ 的分量 x_i 有条件熵 $H(x_i|q)=\frac{1}{2}\sum_{k=1}^{K}h_k\log2\pi e(\boldsymbol{R}_k)_{ii}$, 其中 $(\boldsymbol{R}_k)_{ii}$ 表示矩阵 $\boldsymbol{R}_k$ 的元素 ii .

考虑对随机向量 $\boldsymbol{x}$ 进行一个正交变换 $\boldsymbol{A}$ 得到一个新的随机向量 $\boldsymbol{y}$, 即

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x} \tag{3}$$

[1] 2002-04-22 收到, 2002-08-22 改回
国家自然科学基金资助项目 (批准号: 60073053)

那么 $\boldsymbol{y}$ 的概率密度为 $p(\boldsymbol{y})=\sum_{k=1}^{K}h_kN(\boldsymbol{AR}_k\boldsymbol{A}^T)$. 由 (1) 式，我们有

$$\begin{aligned}\sum_{k=1}^{K}h_k\log\det[\mathrm{diag}(\boldsymbol{AR}_k\boldsymbol{A}^T]&=\sum_{k=1}^{K}h_k\sum_{n=1}^{N}\log(\boldsymbol{AR}_k\boldsymbol{A}^T)_{nn}\\&=\sum_{n=1}^{N}\sum_{k=1}^{K}h_k\log(\boldsymbol{AR}_k\boldsymbol{A}^T)_{nn}=2\sum_{n=1}^{N}H(y_n|q)+c\end{aligned}\tag{4}$$

其中 c 是常数. 所以要使 $\sum_{k=1}^{K}h_k\log\det[\mathrm{diag}(\boldsymbol{AR}_k\boldsymbol{A}^T)]$ 最小，等价于使得随机变量 $\boldsymbol{y}$ 的各个分量的条件熵的和最小. 因为 $\sum_n H(y_n|q)\le\sum_n H(y_n)$ ，这是由于 $\forall n$, $H(y_n|q)\le H(y_n)$ ，所以我们可以寻找正交变换 $\boldsymbol{A}$ 使得 $\boldsymbol{y}$ 的各个分量的熵的和 $\sum_n H(y_n)$ 最小. 注意到 (3) 式和独立分量分析 (ICA) 模型 [3] 的相似性，以及在正交变换 $\boldsymbol{A}$ 下 $\sum_n H(y_n)$ 最小和互信息 $I(\boldsymbol{y})=-\log|\det A|-H(\boldsymbol{x})+\sum H(y_n)$ 最小的等价性，那么我们可以通过 ICA 来近似完成矩阵的联合对角化.

与一般的 ICA 问题不同的是，在这里我们已知的是观察变量 $\boldsymbol{x}$ 的概率分布 $p(\boldsymbol{x})$ 而不是它的样本，并且 $p(\boldsymbol{x})$ 是以混合高斯模型的形式给出的，这些对我们算法的实现都是非常有利的. 如文献 [4] 所述，矩阵 $\boldsymbol{A}$ 的任一行 $\boldsymbol{A}_n$ 可以通过使 $\boldsymbol{y}$ 相应的分量 y_n 的熵 $H(y_n)$ 最小得到. 通过对 $H(y_n)$ 进行近似，本文给出了一种求解 $\boldsymbol{A}_n$ 的牛顿算法，在正交约束之下我们可以得到矩阵 $\boldsymbol{A}$.

2 算　法

由 (3) 式， y_n 的概率密度函数可以表示为混合高斯模型的形式：

$$p(y_n)=\sum_{k=1}^{K}h_kN(\boldsymbol{a}_n^T\boldsymbol{R}_k\boldsymbol{a}_n)\tag{5}$$

其中 $\boldsymbol{a}_n^T$ 是矩阵 $\boldsymbol{A}$ 的第 n 行. 注意到当 q 是离散随机变量时，有 $H(y_n)\le H(q,y_n)$ ，我们可以通过使得 $H(q,y_n)$ 最小来近似使得 $H(y_n)$ 最小. 因为 $H(q,y_n)=H(q)+H(y_n|q)$ ，而 $H(q)$ 是个常数，所以只需 $H(y_n|q)$ 最小就足够了. 对于 $H(y_n|q)$ ，我们有简单的形式：

$$H(y_n|q)=\sum_k h_k\log(\boldsymbol{a}_n^T\boldsymbol{R}_k\boldsymbol{a}_n)+c\tag{6}$$

其中 c 是一个常数. 要寻找使得 $H(y_n|q)$ 最小的 $\boldsymbol{a}_n$ ，我们可以使用梯度下降法，其中梯度为

$$\nabla_{\boldsymbol{a}_n}H(y_n|q)=\sum_k h_k\cdot\frac{2\boldsymbol{R}_k\boldsymbol{a}_n}{\boldsymbol{a}_n^T\boldsymbol{R}_k\boldsymbol{a}_n}\tag{7}$$

若使用牛顿法，由 Kuhn-Tucker 条件，在约束 $\|\boldsymbol{a}_n\|_2=1$ 下，我们有

$$\boldsymbol{F}\equiv\sum_k h_k\cdot\frac{2\boldsymbol{R}_k\boldsymbol{a}_n}{\boldsymbol{a}_n^T\boldsymbol{R}_k\boldsymbol{a}_n}+\beta\boldsymbol{a}_n=0\tag{8}$$

其中 β 是 Lagrange 乘子，并且由上式可以解得 $\beta=-2$. 这样 Hessian 矩阵为

$$\boldsymbol{H}\equiv\nabla_{\boldsymbol{a}_n}^T\boldsymbol{F}=\sum_k h_k\cdot\frac{2(\boldsymbol{a}_n^T\boldsymbol{R}_k\boldsymbol{a}_n)\boldsymbol{R}_k-4\boldsymbol{R}_k\boldsymbol{a}_n\boldsymbol{a}_n^T\boldsymbol{R}_k}{(\boldsymbol{a}_n^T\boldsymbol{R}_k\boldsymbol{a}_n)^2}+\beta\boldsymbol{I}\tag{9}$$

这样我们可以按照下式:

$$a_n = a_n - \alpha H^{-1} F \tag{10}$$

对 a_n 进行搜索，其中 α 是学习速率.

要得到矩阵 A，只需在搜索时对 $a_n, n = 1, \cdots, N$ 进行正交约束即可. 假设我们已经得到了 A 的前 $n-1$ 行 $a_1^T, \cdots, a_{n-1}^T$，那么在按照 (10) 式对 a_n 进行每一步搜索之后，可使用下述 Gram-Schmidt 方法对 a_n 进行正交化.

$$a_n = a_n - \sum_{m=1}^{n-1} (a_n^T a_m) a_m \tag{11}$$

我们也可以同时搜索 A 的所有行直接得到 A，如果这样，那么依照 (10) 式对所有的 $a_n, n = 1, 2, \cdots, N$ 进行迭代，并且在每一步之后进行下面的正交化.

$$A = (AA^T)^{-1/2} A \tag{12}$$

3 实 例

我们对由一组数据使用混合高斯模型学习得到的 12 个 4×4 维的协方差矩阵进行联合对角化，这些数据已经进行了白化. 我们使用 Gram-Schmidt 正交化方法，通过逐一得到矩阵 A 的各个行向量而得到此变换矩阵. 算法的性能如图 1 所示.

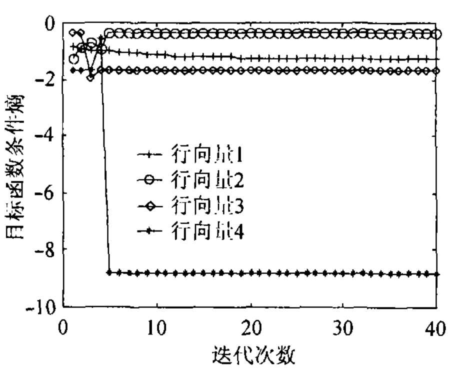

图 1 算法的收敛曲线

其中横轴为迭代次数，纵轴为目标函数条件熵. 我们看到，算法的收敛速度是相当快的. 为了衡量算法的精度，我们计算得到的对角化后矩阵中非对角线元素的均方值 $\frac{1}{N^2-N}\sum_{k=1}^{K} h_k \mathrm{tr}(AR_kA^T - \mathrm{diag}(AR_kA^T))$ 为 4.20×10^{-2}. 在此相同数据集之下，我们与文献 [1] 中的算法进行了比较. 结果表明，直接使用文献 [1] 中算法将陷于局部极小点，误差为 5.13×10^{-2}，如果在迭代中对 A 强制进行正交化，算法将不再收敛.

4 讨 论

在许多应用中，都能够产生和需要解决矩阵的联合对角化问题，例如在文献 [5] 中应用矩阵的联合对角化解决公共主分量分析 (common principal components analysis) 问题. 此外在盲源分离问题中，矩阵的联合对角化也得到了广泛的应用，例如在文献 [7] 的高阶统计量方法中，文献 [6] 中利用源的时间相关特性使用二阶统计量完成盲分离的方法中，以及在文献 [9] 对非平稳信号的分离当中. 已有的矩阵联合对角化算法可见文献 [1,7,8]. 在本文中，我们首先将每一个实对称正定矩阵看作是一个高斯分布随机向量的协方差矩阵，进而引入了一个使用混合高斯模型表示概率密度函数的随机向量. 这样，我们可以将此联合对角化问题近似为一个 ICA 问题，并

给出了相应的算法。如文献 [7] 所述，我们知道 ICA 可以转化为一个联合对角化问题来解决，同样地，由本文的论述，矩阵的联合对角化问题也可以近似为一个 ICA 问题来解决。同时，文中的实验表明了本文提出的算法具有良好的性能。

参 考 文 献

[1] D. T. Pham, Joint approximate diagonalization of positive definite Hermitian matrices, SIAM J. on Matrix Anal. and Appl., 2001, 22(4), 1136–1152.

[2] L. Parra, C. Spence, Separation of Non-Stationary Natural Signals, Independent Components Analysis: Principles and Practice, C. Roberts, *et al.* Ed., Cambridge, U. K.: Cambridge Univ. Press, 2000, 135–157.

[3] P. Comon, Independent component analysis–A new concept? Signal Processing, 1994, 36(3), 287–314.

[4] A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. on Neural Networks, 1999, 10(3), 626–634.

[5] B. Flury, Common principal components in k groups, J. Am. Statist. Assoc., 1984, 79(5), 892–897.

[6] A. Belouchrami, K. Abed-Meraim, J. Cardoso, E. Moulines, A blind source separation technique using second-order statistics, IEEE Trans. on Signal Processing, 1997, 45(2), 434–444.

[7] J. Cardoso, A. Souloumiac, Blind beamforming for non-Gaussian signals, IEE Proc.-F, 1993, 140(6), 362–370.

[8] B. Flury, W. Gautschi, An algorithm for the simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly orthogonal form, SIAM J. of Sci. Statist. Comp., 1986, 7(1), 169–184.

[9] D.T. Pham, J. Cardoso, Blind separation of instantaneous mixtures of nonstationary sources, IEEE Trans. on Signal Processing, 2001, SP-49(9), 1837–1848.

AN ALGORITHM FOR JOINT DIAGONALIZATION OF POSITIVE DEFINITE SYMMETRIC MATRICES

Yue Bo Jiao Licheng

(*Key Lab. for Radar Signal Processing, Xidian University, Xi'an* 710071, *China*)

Abstract This paper proposes a new algorithm for joint diagonalization of positive definite symmetric matrices and shows that it can be approximated by an independent component analysis problem. The good performance of the algorithm is illustrated with an example by compared with other approaches.

Key words Joint diagonalization, Independent Component Analysis(ICA), Gaussian distribution

岳 博： 男，1970 年生，博士生，主要研究方向为概率模型，Bayes 统计方法等.

焦李成： 男，1959 年生，教授，博士生导师，主要研究领域有非线性科学，智能信息处理，神经网络，数据挖掘等.