

# 自适应拟牛顿子空间跟踪算法分析<sup>1</sup>

马长征 吴伟陵

(北京邮电大学信息工程学院 北京 100876)

**摘要** 该文对自适应拟牛顿子空间跟踪算法进行了详细分析,证明了其本质上类似于幂法,但一般来说暂态性能劣于幂法,给出了在相邻数据之间只做一次迭代时特征值、特征向量估计的误差,计算机仿真结果证明了该文分析的正确性。

**关键词** 子空间跟踪, 拟牛顿法, 特征值, 特征向量

**中图分类号** TN911.7, TN911.23

## 1 引言

在雷达、通信等信号处理中估计白噪声中窄带信号的参数是一个重要的研究课题。其中的一类重要方法就是特征结构法,即利用信号协方差矩阵的特征值和特征向量来对参数进行估计。如阵列信号处理中的 MUSIC, ESPRIT 等方法, CDMA 盲多用户检测中基于特征结构的方法。一般情况下信号环境是时变的,信号协方差矩阵的特征值和特征向量也是时变的,就需要对特征空间进行跟踪。已有大量文献讨论子空间的跟踪<sup>[1-5]</sup>。B. Yang 在文献 [1] 中提出了一种收缩投影近似子空间跟踪算法,其复杂度是  $O(nr)$ ,其中  $n$  是协方差矩阵的维数,  $r$  是需要估计的特征向量的个数。Mathew 等在文献 [2] 中利用神经网络的方法极小化一个代价函数,给出了非负定矩阵的小特征值对应的特征向量构成的特征空间的估计,在此基础上, Mathew 等又在文献 [3] 中提出拟牛顿法及填充技术估计全部或部分特征值对应的特征向量。Kang 等利用收缩技术及拟牛顿法估计大特征值对应的特征向量<sup>[4]</sup>,但 Kang 对 Hesse 矩阵的逆的近似比 Mathew 的精确。文献 [3, 4] 的复杂度是  $O(n^2r)$ ,但文献 [4] 的仿真结果表明,文献 [4] 的方法的收敛性能和跟踪性能比文献 [1, 3] 都要好。本文通过分析发现文献 [3, 4] 的拟牛顿法由于对 Hesse 矩阵取近似,本质上都是一种幂法,性能不比传统的幂法好。另外给出了幂法一步迭代的误差。仿真结果表明,在相连的两个采样点之间,只需要一次迭代。

## 2 自适应拟牛顿算法

设  $x_k \in \mathbb{R}^N$  是广义平稳或慢时变信号的快拍矢量,  $R$  是数据的协方差矩阵,我们的目的是求  $R$  的前  $r$  个大特征值及对应的特征向量。在实际情况下一般都满足这样的假设,  $R$  是对称正定的,特征值  $\lambda_i$  满足  $\lambda_1 > \lambda_2 > \dots > \lambda_r > \lambda_{r+1} \geq \dots \geq \lambda_N > 0$ 。求  $\lambda_1$  对应的特征向量  $w$  转变为如下的极小化问题

$$\begin{aligned} \min J(w) &= -w^T R w \\ \text{s.t. } w^T w &= 1 \end{aligned} \quad (1)$$

约束  $w$  的范数为 1,避免无穷大平凡解。设  $\mu$  为某正数,通过罚函数的方法,上述带约束的极小化问题可以转化为下面没有约束的极小化问题。

<sup>1</sup> 2001-03-15 收到, 2001-08-13 改回

$$\min J(w) = -w^T R w + \mu(w^T w - 1)^2 \quad (2)$$

求最小化问题的解可以利用最速下降法, 亦可以利用牛顿法等.  $J(w)$  的梯度为

$$g = \nabla J = -2R w + 4\mu(w^T w - 1)w \quad (3)$$

Hesse 矩阵为

$$H = -2R + 8\mu w w^T + 4\mu(w^T w - 1)I_N \quad (4)$$

牛顿法的迭代格式是

$$w_{k+1} = w_k - H_k^{-1} g_k \quad (5)$$

直接求  $H_k$  的逆计算量是  $O(N^3)$ , 为了节约计算量, 文献 [3] 略去了  $4\mu(w^T w - 1)I_N$ . 由文献 [3, 4] 的分析, 当  $k \rightarrow \infty$  时,  $4\mu(w_k^T w_k - 1)I_N \rightarrow 2\lambda_1$ ,  $4\mu(w_k^T w_k - 1)I_N$  是和  $2R$  同等量级的量, 所以, 略去了  $4\mu(w_k^T w_k - 1)I_N$  将引入较大的误差. 令  $Q = 2\mu(w^T w - 1)I_N - R$ , 文献 [4] 利用了一个近似求逆公式,  $(I - P)^{-1} \approx I + P$ , 得到  $Q^{-1} \approx \frac{1}{2\mu(w^T w - 1)} \left[ I_N + \frac{R}{2\mu(w^T w - 1)} \right]$ . 我们知道, 只有当  $\|P\| \ll 1$  时,  $(I - P)^{-1} \approx I + P$  才成立, 但由前面的分析, 这一条件是不成立的, 所以,  $Q^{-1}$  的表达式也存在较大误差.  $2Q + 8\mu w w^T$  的逆可以由秩一修正矩阵求逆公式精确求出. 文献 [4] 最终的迭代公式为

$$w_{k+1} = L_k w_k \quad (6)$$

$$L_k = \frac{w_k^T w_k}{\tau_k} [2\mu(w_k^T w_k - 1)I_N + R] \quad (7)$$

$$\tau_k = \mu(w_k^T w_k - 1)^2 + 2\mu w_k^T w_k (w_k^T w_k - 1) + w_k^T R w_k \quad (8)$$

文献 [4] 证明了当初始向量充分接近真正的特征向量时, 上述格式收敛于真值. 实际的初始值不一定接近真正的特征向量, 上述格式可能出问题. 如令  $w_0 = 0$ , 则  $w_k \rightarrow 0$ , 收敛到了错误的值. 在我们的仿真中也经常出现收敛到 0 的情况.

对于其他的特征值和特征向量可以利用收缩技术. 设  $w^1$  是  $\lambda_1$  对应的特征向量, 构造  $R^1 = R - w^1 w^{1T} R$ , 使得  $R^1$  中不含  $w^1$  的分量, 那么  $R^1$  的最大特征值为  $\lambda_2$ , 利用 (6-8) 式也可以进行迭代求  $\lambda_2$  和  $w^2$ , 依次类推.

由 (7) 式可以看到,  $L_k$  是  $R$  的线性函数. 下面我们证明, 任何  $L_k$  可由  $R$  的系数为正的  $M$  阶多项式表示. 迭代格式如 (6) 式的方法, 其收敛性能都不如简单的幂法, 即令  $L_k = R^M$  的性能. 设  $L_k = \sum_{j=0}^M \beta_j R^j$ ,  $\beta_j > 0$ ,  $\{w^j\}_{j=1}^N$  为  $R$  的归一化特征向量, 初始迭代向量  $w_0 = \sum_{i=1}^N \alpha_i w^i$ , 那么

$$L_k w_0 = \left[ \sum_{j=0}^M \beta_j R^j \right] \left[ \sum_{i=1}^N \alpha_i w^i \right] = \sum_{i=1}^N \alpha_i \left[ \sum_{j=0}^M \beta_j \lambda_i^j \right] w^i \quad (9)$$

迭代的收敛性能可由  $w^1$  前面的系数和  $\{w^i\}$ ,  $i \neq 1$  前面的系数之比衡量. 由  $\beta_j > 0$ ,  $\lambda_1 > \lambda_2 > \dots \geq \lambda_N > 0$ , 易证

$$\frac{\sum_{j=0}^M \beta_j \lambda_1^j}{\sum_{j=0}^M \beta_j \lambda_i^j} < \frac{\lambda_1^M}{\lambda_i^M} \quad (10)$$

即  $L_k = R^M$  的收敛性能比  $L_k = \sum_{j=0}^M \beta_j R^j$ ,  $\beta_j > 0$  的收敛性能好。对于 (7) 式, 由于  $2\mu(w_k^T w_k - 1)I_N \rightarrow \lambda_1$ , 故一般满足  $\beta_j > 0$  的要求, 所以还不如令  $L_k = R$ , 迭代公式就变成了简单的幂迭代公式。通过简单的推导, (5) 式还可以写为

$$w_{k+1} = H_k^{-1} 8\mu(w_k^T w_k) w_k \quad (11)$$

(11) 式类似于常规的带位移的反幂加速迭代法<sup>[6]</sup>, 在收敛点附近, 反幂加速迭代法的收敛速度是很高的, 一般用于处理特征值之间距离很近、幂法收敛较慢的情形, 但需要精确的矩阵求逆。

### 3 幂法误差分析

设  $x_k \in \mathbb{R}^N$  是慢时变信号的快拍矢量,  $k$  时刻数据的协方差矩阵  $R_k$  可由如下的秩一方程修改。

$$R_k = \alpha R_{k-1} + (1 - \alpha) x_k x_k^T \quad (12)$$

其中  $\alpha$  是接近 1 的一个数。对广义平稳信号,  $R_k$  可由:  $R_k = [(k-1)/k]R_{k-1} + (1/k)x_k x_k^T$  修改。 $(1 - \alpha)x_k x_k^T$  可以看成是对  $\alpha R_{k-1}$  的一个扰动, 由矩阵特征值的扰动理论,  $R_k$  和  $\alpha R_{k-1}$  的特征值和特征向量之间的关系可由如下的引理描述:

**引理 1**<sup>[6]</sup> 设  $A' = A + \delta$ ,  $\|\delta\|_2 = \varepsilon$  是一个小量,  $\lambda$  和  $\varphi$  是  $A$  的特征值和归一化特征向量, 其它特征值记为  $\lambda_i$ , 特征向量组成的矩阵记为  $Q$ , 令  $S = Q \text{diag}(\lambda_i - \lambda)^{-1} Q^T$ , 那么, 存在  $A'$  的特征值和归一化特征向量  $\lambda'$  和  $\varphi'$ , 满足

$$\lambda' = \lambda + \varphi^T \delta \varphi + O(\varepsilon^2) \quad (13)$$

$$\varphi' = \varphi - S \delta \varphi + O(\varepsilon^2) \quad (14)$$

设  $\lambda_{k-1}^1$  和  $\varphi_{k-1}^1$  分别是  $R_{k-1}$  的特征值和特征向量, 那么, 存在  $R_k$  的特征值和特征向量  $\lambda_k^1$  和  $\varphi_k^1$ , 其与  $\lambda_{k-1}^1, \varphi_{k-1}^1$  的关系满足 (13), (14) 式。令  $R_k$  乘  $\varphi_{k-1}^1$  得

$$\hat{\varphi}_k^1 = R_k \varphi_{k-1}^1 = \alpha \lambda_{k-1}^1 \varphi_{k-1}^1 + (1 - \alpha) x_k^T \varphi_{k-1}^1 x_k \quad (15)$$

$$\begin{aligned} \hat{\varphi}_k^1 - \alpha \lambda_{k-1}^1 \varphi_k^1 &= (1 - \alpha) x_k^T \varphi_{k-1}^1 x_k + (1 - \alpha) x_k^T \varphi_{k-1}^1 \alpha \lambda_{k-1}^1 S x_k \\ &= (1 - \alpha) x_k^T \varphi_{k-1}^1 [I + \alpha \lambda_{k-1}^1 S] x_k \end{aligned} \quad (16)$$

归一化的  $\hat{\varphi}_k^1$  与  $\varphi_k^1$  的差为  $(1 - \alpha)O(1)$ 。

$$\hat{\lambda}_k^1 = \varphi_{k-1}^1{}^T R_k \varphi_{k-1}^1 = \alpha \lambda_{k-1}^1 + (1 - \alpha) \|x_k^T \varphi_{k-1}^1\|^2 \quad (17)$$

$$\hat{\lambda}_k^1 - \lambda_k^1 = O(\varepsilon^2) \quad (18)$$

由 (16), (18) 式可以看到, 由  $R_{k-1}$  的精确特征向量作为初始迭代向量, 一步迭代后特征向量的精度为  $(1-\alpha)O(1)$ , 特征值的精度为  $O(\varepsilon^2)$ , 若以  $\hat{\varphi}_k^1$  代替  $\varphi_{k-1}^1$  做 Rayleigh 商, 特征值应该有更高的精度。实际上不可能得到  $R_{k-1}$  的精确的特征向量, 经与上面类似的推导可知, 即使从与  $\varphi_{k-1}^1$  有  $(1-\alpha)O(1)$  误差的特征向量出发, 一次迭代得到的归一化的  $\hat{\varphi}_k^1$  与  $\varphi_k^1$  的差也为  $(1-\alpha)O(1)$ , 特征值的差也为  $(1-\alpha)O(1)$ 。

直接利用子空间迭代, 复杂度为  $O(n^2r) + O(nr^2)$ , 与拟牛顿法一样, 性能也相当稳健。设  $Q_0$  是  $r$  个正交向量组成的矩阵, 子空间迭代格式如下:

- (1) 计算  $Y_k = R_k Q_{k-1}$ ;
- (2) 对  $Y_k$  进行 QR 分解, 即  $Q_k A_k = Y_k$ 。

#### 4 仿真结果

由于文献 [1, 3, 4] 方法的性能已经在文献 [4] 中进行了比较, 我们在这里只比较拟牛顿法 (NT), 幂法 (PT), 子空间法 (ST)。我们首先对平稳信号评估各种方法的性能, 子空间跟踪性能指标按照文献 [5] 给出的子空间信噪比 (SSNR)。设  $Q$  是估计出来的特征向量组成的矩阵,  $V$  是真正的特征向量组成的矩阵, 令  $\theta = V^T Q$ , 定义:

$$\text{SSNR} = 10 \log_{10} \left[ \frac{\text{tr}(\theta^T \theta)}{\text{tr}(I - \theta^T \theta)} \right] \quad (19)$$

协方差矩阵  $R$  的特征值与文献 [4] 中的一样取为: 11.7996, 5.5644, 3.4175, 2.0589, 0.7873, 0.5878, 0.1743, 0.1423, 0.1213, 0.1007, 有 4 个大特征值, 特征向量取为任意的一组正交向量  $V$ , 令  $R = V \Lambda V^T = (V \Lambda^{1/2})(\Lambda^{1/2} V) = U U^T$ , 零均值的高斯随机向量过程  $\{w(t)\}$  满足  $E\{w w^T\} = I$ , 仿真数据序列由  $x(t) = U w(t)$  产生。初始迭代矩阵为任意正交矩阵, 做 100 次随机实验进行平均。NT 方法中取  $\mu = 10$ , 若某特征向量收敛到 0, 则不计算此次实验, 100 次中约有 5 次收敛到 0。在图 1 中,  $\circ$  代表  $R$  理想的特征子空间和每步用 Matlab 精确计算  $R_k$  的特征子空间的 SSNR, 其他为各种迭代方法估计的  $R_k$  的特征子空间和用 Matlab 精确计算  $R_k$  的特征子空间的 SSNR。ST, NT, PT 分别代表子空间法、Newton 法和幂法。可以看到, 3 种方法的稳态性能一样, NT 的暂态性能较差, 因为前面已经证明拟牛顿法不如幂法。由于数据有限,  $R_k$  的估计的特征子空间和精确子空间的 SSNR 比  $R_k$  的精确子空间和  $R$  的子空间的 SSNR 高很多, 说明数据序列之间只需要一步迭代。图 2 中, 计算出的是第 4 个特征值的误差  $e$  的精度位数, 由  $\log_{10} e$  算出。可以看到 3 种方法对特征值的估计稳态性能一样, 只需要一步迭代。

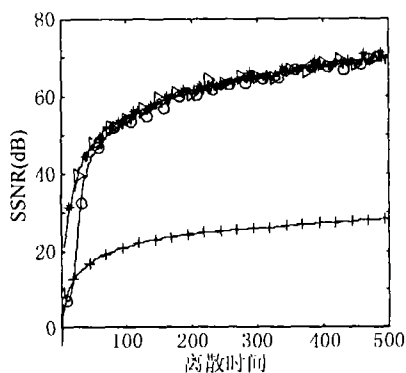


图 1 平均 SSNR

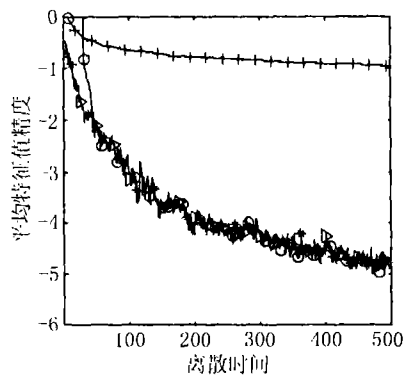


图 2 特征值估计精度

图 3, 图 4 为时变环境下 100 次平均子空间跟踪的情况。在 0 时刻的特征值同例一样, 在 1000 时刻, 在原始特征值上加  $[0, 0.5]$  之间的均匀随机变量。在 0 时刻的特征向量为随机正交向量, 在 1000 时刻的特征向量也为随机正交向量。中间时刻的特征值为线性变化, 特征向量也为线性变化, 但由 QR 分解保持正交性。从图 3、图 4 可以看到, 由于时变性,  $R$  和  $R_k$  之间特征值与特征向量的差变得较大, 自适应算法对  $R_k$  的特征值和特征向量的跟踪精度仍较高。

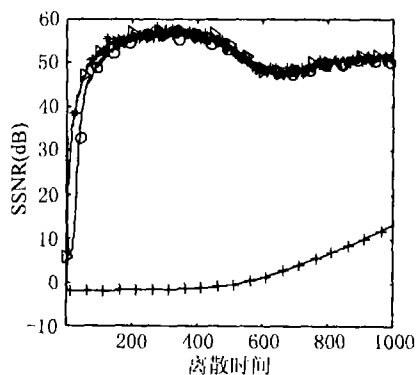


图 3 特征值估计精度

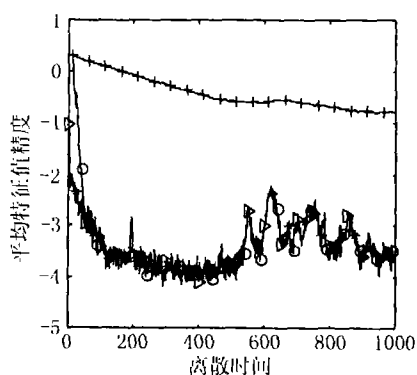


图 4 平均 SSNR

## 5 结 论

本文首先分析了拟牛顿自适应子空间跟踪算法, 证明了在一般情况下其性能劣于幂法。另外, 拟牛顿法并非对任意初始点都收敛到真值, 所以实用性受到了一定的限制。对于幂法, 我们给出了一步迭代的误差, 证明自适应算法在相邻数据之间只需要一次迭代。

## 参 考 文 献

- [1] Yang Bin, Projection approximation subspace tracking, IEEE Trans. on SP, 1995, (43)1, 95-107.
- [2] G. Mathew, V. U. Reddy, Orthogonal eigensubspace estimation using neural networks, IEEE Trans. on SP, 1994, (42)7, 1803-1811.

- [3] G. Mathew, V. U. Reddy, S. Dasgupta, Adaptive estimation of eigensubspace, IEEE Trans. on SP, 1995, (43)2, 401-411.
- [4] Z. J. Kang, *et al*, An adaptive quasi-Newton algorithm for eigensubspace estimation, IEEE. Trans. on SP, 2000, (48)12, 3328-3333.
- [5] P. Strobach, Square-root QR inverse iteration for tracking the minor subspace, IEEE Trans. on SP, 2000, (48)11, 2994-2999.
- [6] F. Chatelin, Spectral Approximation of Linear Operators, New York, Academic Press Inc, 1983, 1-84.

## ANALYSIS OF ADAPTIVE QUASI-NEWTON ALGORITHM FOR EIGENSUBSPACE TRACKING

Ma Changzheng     Wu Weiling

*(Info. Eng. School, Beijing Univ. of Posts and Telecom., Beijing 100876, China)*

**Abstract** In this paper, an adaptive quasi-Newton algorithm for eigensubspace estimation is analysed and proved. This algorithm is similar to power method, but its transient performance is poor than power method. Assuming one iteration to be done between snapshots, the error between estimated and true eigenpairs is obtained. The simulation results show the correct of the analysis in this paper.

**Key words** Subspace tracking, Quasi-Newton, Eigenvalue, Eigenvector

马长征: 男, 1967 年生, 博士后, 研究方向: 智能天线、多用户检测、雷达成像。

吴伟陵: 男, 1938 年生, 教授, 博士生导师, 国家自然科学基金委学科评审组成员, 中国电子学会信息论分会主任委员, 研究方向: 数字信号处理, 个人通信与数字卫星通信中的信息处理, 信源与信道编码理论。