

非最小相位 FIR 系统的自适应辨识¹⁾

宋 宇 张贤达 李衍达

(清华大学自动化系 北京 100084)

摘 要 提出了一种基于线性代数方程组约束和梯度法的非最小相位 FIR 系统的自适应辨识算法. 选用的线性方程组具有列满秩的系数矩阵, 保证了系统参数的唯一可识别性. 由于只采用高阶累量, 故能够抑制任何高斯有色噪声的影响. 重点讨论了梯度法中步长的选择, 提出了收敛速度最快的变步长. 仿真实验的结果证实了算法的有效性.

关键词 自适应系统辨识, 非最小相位系统, 高阶累量.

1 引言

信号处理的许多应用领域涉及非最小相位系统的辨识和反卷积, 如地震勘探、通信、雷达、声呐、图像处理等. 近年来, 基于高阶累量或高阶谱的非最小相位 FIR 系统的辨识方法受到了重视^[1,2]. Mendel 把这些方法分为三类^[1]: 闭式解、线性代数解和非线性优化解, 其中线性代数解法因为计算简单、没有局部极值点和鲁棒性强等原因而广泛使用.

基于观测信号的相关函数即二阶累量的自适应系统辨识理论一直受到重视, 已经提出了 RLS 算法、辅助变量法、预测误差法、贝叶斯法和随机逼近法等诸多算法^[3]. 但它们不可能辨识非最小相位系统, 因为自相关函数不包含相位信息^[4]. 基于高阶累量的自适应算法目前较少. 文献[5]提出了一种同时使用相关函数和高阶累量的超定递推辅助变量 (ORIV) 算法来实现 FIR 和 IIR 系统的自适应辨识. 由于使用了相关函数, ORIV 算法对高斯有色噪声是敏感的, 同时它需要将累量构成的线性方程组的系数矩阵分解成两个特殊矩阵的乘积, 故不能保证适用于系统辨识的所有线性代数解法. 例如, 文献[2]中的方法就无法做这样的分解. 文献[6]通过用非因果 AR 系统模拟非最小相位 FIR 系统的逆系统, 采用梯度法实现了基于高阶累量的 FIR 系统的自适应反卷积. 梯度法计算简单, 但收敛较慢. 由于高阶累量的估计性能一般比相关函数差, 基于高阶累量的自适应算法比基于相关函数的收敛更慢. 加快其收敛速度是一个重要的问题.

本文结合非线性优化理论中简单而有效的梯度法和文献[2]提出的 FIR 系统辨识的线性代数解法, 设计了基于高阶累量的自适应系统辨识算法. 为了改进算法的收敛性能, 重点分析了最佳变步长的选择, 放宽了现有文献对梯度型自适应算法中步长取值范围的限制.

1) 国家自然科学基金资助项目.
收稿日期: 1994-05-03

2 FIR 系统辨识的线性代数方法

设观测信号为一未知 FIR 系统的输出

$$x(n) = \sum_{i=0}^q b(i)w(n-i) + v(n). \quad (1)$$

其中输入信号 $w(n)$ 是一个零均值、独立同分布 (*i. i. d.*) 的非高斯随机过程. $b(i)$ 是 FIR 系统的参数, 不失一般性, 令 $b(0)=1$, $b(q) \neq 0$, 且系统为非最小相位. 加性噪声 $v(n)$ 是一个高斯 ARMA 过程, 与输入 $w(n)$ 独立, 从而也与输出 $x(n)$ 独立. 问题是在 $b(i)$, $w(n)$ 和 $v(n)$ 均未知的情况下, 随着观测信号 $x(n)$ 的来自适应地估计参数 $b(i)$, 从而辨识此 FIR 系统. 这是一个盲辨识问题.

FIR 系统辨识的线性代数方法就是求解约束 FIR 系统的线性代数方程

$$A\theta = c. \quad (2)$$

其中 A 和 c 分别为观测信号 $x(n)$ 的累量构成的矩阵和向量, θ 为待求的 FIR 系统参数构成的向量. 文献[2]提出了两种只用高阶累量辨识非最小相位 FIR 系统的线性代数方法. 以第一种方法为例, 式(2)中的矩阵、向量和最后的参数估计值分别为

$$A = \begin{bmatrix} C_{mx}^{m-1}(q, q) & & & 0 \\ C_{mx}^{m-1}(q, q-1) & C_{mx}^{m-1}(q, q) & & \\ \vdots & \vdots & \ddots & \\ C_{mx}^{m-1}(q, 0) & C_{mx}^{m-1}(q, 1) & \cdots & C_{mx}^{m-1}(q, 0) \\ & C_{mx}^{m-1}(q, 0) & \cdots & C_{mx}^{m-1}(q-1) \\ & & \ddots & \vdots \\ 0 & & & C_{mx}^{m-1}(q, 0) \end{bmatrix} \quad (3a)$$

$$c = C_{mx}^{m-3}(q, 0)C_{mx}(q, q) \begin{bmatrix} C_{mx}(-1, 0) \\ C_{mx}(-q+1, 0) \\ \vdots \\ C_{mx}(0, 0) \\ C_{mx}(1, 0) \\ \vdots \\ C_{mx}(q, 0) \end{bmatrix} \quad (3b)$$

$$\theta = [b(0), b(1), \dots, b(q)]^T, \quad (3c)$$

$$\hat{b}(i) = \frac{b(i)}{b(0)}, \quad i = 1, \dots, q. \quad (3d)$$

式中 $C_{mx}(i, j) = C_{mx}(i, j, 0, \dots, 0)$ 表示信号 $x(n)$ 的 m 阶累量. 可以看出方程组的个数为 $2q+1$, 而未知数的个数是 $q+1$, 所以这是一个超定方程组. 已经证明^[2], 矩阵 A 是列满秩的, 因此方程组具有唯一的最小二乘解

$$\theta^* = Q^{-1}A^T c, \quad Q = A^T A. \quad (4), (5)$$

即是说, MA 参数是唯一可辨识的. 此外, 由于方程组仅由高阶累量构成(在式(3)中取 $m > 2$), 而高斯随机过程的高阶累量恒等于零, 所以求解的结果理论上完全不受加性高斯有

色噪声的影响.

3 自适应算法

3.1 累量的递推更新

根据第 2 节的辨识方法,只要能递推地计算矩阵 A 和向量 c ,就能按照梯度法设计估计 θ 的自适应算法.这就需要先得到累量的递推公式.根据定义^[1],容易推出二至四阶累量的递推公式(k 为时间)如下:

$$\hat{C}(i)_{2r}^{(k)}(m) = \frac{k-1}{k} \hat{C}(i)_{2r}^{(k-1)}(m) + \frac{1}{k} x(k)x(k-|m|), \quad (6)$$

$$\begin{aligned} \hat{C}(i)_{3r}^{(k)}(m,n) &= \frac{k-1}{k} \hat{C}(i)_{3r}^{(k-1)}(m,n) + \frac{1}{k} x(k-i)x(k-i+m)x(k-i+n), \\ i &= \max(0,m,n), \end{aligned} \quad (7)$$

$$\hat{C}(i)_{4r}^{(k)}(l,m,n) = I_1^{(k)}(l,m,n) - I_2^{(k)}(l,m,n), \quad (8a)$$

$$\begin{aligned} I_1^{(k)}(l,m,n) &= \frac{k-1}{k} I_1^{(k-1)}(l,m,n) + \frac{1}{k} x(k-j)x(k-j+l)x(k-j+m) \\ &\quad \times x(k-j+n), \quad j = \max(0,l,m,n), \end{aligned} \quad (8b)$$

$$\begin{aligned} I_2^{(k)}(l,m,n) &= \hat{C}(i)_{2r}^{(k)}(l) \hat{C}(i)_{2r}^{(k)}(m-n) + \hat{C}(i)_{2r}^{(k)}(m) \hat{C}(i)_{2r}^{(k)}(l-n) \\ &\quad + \hat{C}(i)_{2r}^{(k)}(n) \hat{C}(i)_{2r}^{(k)}(l-m). \end{aligned} \quad (8c)$$

注意四阶累量的递推更新需辅之以二阶累量的递推更新.式(6)~(8)是观测信号为平稳随机过程时的递推公式.如果系统参数是时变的,观测信号 $x(n)$ 是非平稳随机过程.此时累量的递推应取如下形式:

$$C^{(k)} = \lambda C^{(k-1)} + \Delta^{(k)}, \quad (9)$$

其中 Δ 是修正项,而 λ 是遗忘因子

$$0 \leq \lambda \leq 1. \quad (10)$$

λ 的选取决定于跟踪时变系统的能力和抑制噪声影响之间的平衡^[3].

3.2 参数的自适应估计

如果已知 $k-1$ 时刻的累量估计值和 k 时刻的观测信号,就可以由式(6)~(8)得到 k 时刻的累量估计值,即可递推地计算方程组(2)的系数矩阵和向量,从而得到时变的线性方程组

$$\hat{A}_k \hat{\theta}_k = \hat{c}_k. \quad (11)$$

令 k 时刻的误差向量为

$$\hat{e}_k = \hat{c}_k - \hat{A}_k \hat{\theta}_k, \quad (12)$$

并定义 k 时刻的性能函数为

$$\hat{\xi}_k(\hat{\theta}_k) = \frac{1}{2} \|\hat{e}_k\|^2 = \frac{1}{2} \hat{\theta}_k^T \hat{Q}_k \hat{\theta}_k - \hat{\theta}_k^T \hat{A}_k^T \hat{c}_k + \frac{1}{2} \hat{c}_k^T \hat{c}_k, \quad (13)$$

其中

$$\hat{Q}_k = \hat{A}_k^T \hat{A}_k. \quad (14)$$

显然, $\hat{\xi}_k$ 的负梯度是

$$\hat{\mathbf{g}}_k = - \frac{\partial \hat{\xi}_k(\hat{\boldsymbol{\theta}}_k)}{\partial \hat{\boldsymbol{\theta}}_k} = \hat{\mathbf{A}}_k^T \hat{\mathbf{e}}_k. \quad (15)$$

用梯度法搜索此二次型的性能表面,即得到参数的自适应估计

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \mu_k \hat{\mathbf{g}}_k, \quad (16)$$

式中 μ_k 是可变步长.

3.3 步长的选择

梯度型自适应算法中步长的选择与算法的收敛性能关系密切. 现有文献对梯度型自适应算法中步长的基本要求是^[7]

$$0 < \mu < \frac{2}{\lambda_{\max}}. \quad (17)$$

其中 λ_{\max} 为自相关矩阵 R (对应于本文式(5)中的 Q) 的最大特征值. 算法的收敛速度取决于 R 的特征值的分布范围, 其分布越广, 算法收敛越慢. 实际上式(17)是采用恒定步长时步长的收敛范围. 如果允许步长可变, 以上约束可以放宽. 在确定性性能函数的梯度法搜索中, 最佳步长是^[8]

$$\mu_k = (\mathbf{g}_k^T \mathbf{g}_k) / (\mathbf{g}_k^T \mathbf{Q}_k \mathbf{g}_k). \quad (18)$$

在自适应算法(16)中, 取步长

$$\mu_k = (\hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k) / (\hat{\mathbf{g}}_k^T \hat{\mathbf{Q}}_k \hat{\mathbf{g}}_k), \quad (19)$$

它是梯度的瑞利商的倒数.

$$\text{令 } \Delta \hat{\xi}_k = \xi_k(\hat{\boldsymbol{\theta}}_{k+1}) - \hat{\xi}_k(\hat{\boldsymbol{\theta}}_k), \quad (20)$$

它是 k 时刻用梯度法搜索后的性能函数的增量. 利用式(12)~(16), (19), (20), 容易证明

$$(\partial \Delta \hat{\xi}_k) / (\partial \mu_k) = 0. \quad (21)$$

由于是在负梯度方向搜索, 所以 $\Delta \hat{\xi}_k < 0$. 式(21)说明 μ_k 的选择使参数估计 $\hat{\boldsymbol{\theta}}_k$ 在性能函数 $\hat{\xi}_k$ 的负梯度方向上前进一段最佳的距离而得到最接近极值点的新的估计 $\hat{\boldsymbol{\theta}}_{k+1}$. 采用式(19)定义的步长可使自适应算法(16)具有最快的收敛速度. 这种最佳步长是有界的^[9]

$$\frac{1}{\hat{\lambda}_{\max}^{(k)}} \leq \mu_k \leq \frac{1}{\hat{\lambda}_{\min}^{(k)}}, \quad (22)$$

式中 $\hat{\lambda}_{\max}^{(k)}$ 和 $\hat{\lambda}_{\min}^{(k)}$ 分别为 $\hat{\mathbf{Q}}_k$ 的最大和最小特征值. 从式(19)容易看出 μ_k 仅与 $\hat{\mathbf{g}}_k$ 的方向有关. 实际上, 在递推过程中负梯度 $\hat{\mathbf{g}}_k$ 的方向一直在改变. 当 $\hat{\mathbf{g}}_k$ 是对应于 $\hat{\lambda}_{\max}^{(k)}$ 的特征向量时, μ_k 取最小值 $1/\hat{\lambda}_{\max}^{(k)}$; 当 $\hat{\mathbf{g}}_k$ 是对应于 $\hat{\lambda}_{\min}^{(k)}$ 的特征向量时, μ_k 取最大值 $1/\hat{\lambda}_{\min}^{(k)}$; 一般情况下, μ_k 则介于最大和最小值之间. 因此, 式(19)的几何意义是根据梯度的方向自动调节步长. 综合以上分析结果, 本文提出的算法(采用三阶累量辨识时不变系统的情况)见表 1. 采用四阶累量或辨识时变系统时只需更改累量递推步骤即可. 表 1 的算法每步迭代需 $6q^2 + 28q + 13$ 次乘法, 优于 ORIV 算法^[5] 的 $42q^2 + 71q + 24$. 如果按照文献[6]中的方法计算步长

$$\mu_k = 0.9 / \text{tr} \hat{\mathbf{Q}}_k, \quad (23)$$

则需要对对称矩阵 $\hat{\mathbf{Q}}_k$ 作特征分解, 而典型的特征分解算法^[10] 的计算量约为 $\frac{2}{3}(g+1)^3$. 相比之下, 利用式(19)计算步长只需 $2q^2 + 6q + 3$ 次乘法.

表 1 自适应算法总结

Initialization	Dimension
$\hat{C}_{3,r}^{(0)} = 0$	$-q : q \times 0 : q$
$\hat{\theta}_1 = 0$	$(q+1) \times 1$
For $k=1, 2, \dots$	
For $m=-q, \dots, q$	
$i = \max(0, m)$	
$\hat{C}_{3,r}^{(k)}(m, 0) = \frac{k-1}{k} \hat{C}_{3,r}^{(k-1)}(m, 0) + \frac{1}{k} x^2(k-i)x(k-i+m)$	$-q : q \times 0 : q$
For $n=0, \dots, q$	
$\hat{C}_{3,r}^{(k)}(q, n) = \frac{k-1}{k} \hat{C}_{3,r}^{(k-1)}(q, n) + \frac{1}{k} x(k-q)x(k-q+n)x(k)$	$-q : q \times 0 : q$
Form \hat{A}_k and \hat{c}_k	
$\hat{e}_k = \hat{c}_k - \hat{A}_k \hat{\theta}_k$	$(2q+1) \times 1$
$\hat{g}_k = \hat{A}_k^T \hat{e}_k$	$(q+1) \times 1$
$h_k = \hat{A}_k \hat{g}_k$	$(2q+1) \times 1$
$\mu_k = \frac{\hat{g}_k^T \hat{g}_k}{h_k^T h_k}$	
$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu_k \hat{g}_k$	$(q+1) \times 1$
For $i=1, \dots, q$	
$\hat{b}_{k+1}(i) = \frac{\hat{\theta}_{k+1}(i+1)}{\hat{\theta}_{k+1}(1)}$	

4 实验结果

此处给出两个典型的低信噪比仿真实验结果. 所有实验中加入性观测噪声 $v(n)$ 是高斯 ARMA(3, 1) 过程^[2]:

$$v(n) = 2.2v(n-1) + 1.77v(n-2) - 0.52v(n-3) = e(n) - 1.25e(n-1).$$

输入信号 $w(n)$ 是指数分布随机过程 ($\gamma_1=0, \gamma_2=1, \gamma_3=2$). 数据长度为 5000. 每例分别做 20 次 Monte Carlo 试验. 实验中假定模型阶数已知.

例 1. (时不变系统的辨识) 真实非最小相位 MA 模型是^[2]

$$x(n) = w(n) + 0.9w(n-1) + 0.385w(n-2) - 0.771w(n-3) + v(n).$$

它的零点位于 $-0.75 \pm j0.85$ 和 0.6 . 信噪比 SNR = 5 dB. 图 1 是采用表 1 的算法得到的参数的自适应数据估计. 从图中可以看出大约经过 700 个数据点后估计值趋于收敛. 经过 5000 个点后参数的估计值分别为 $\hat{b}(1) = 0.929 \pm 0.143, \hat{b}(2) = 0.424 \pm 0.227, \hat{b}(3) = -0.808 \pm 0.175$.

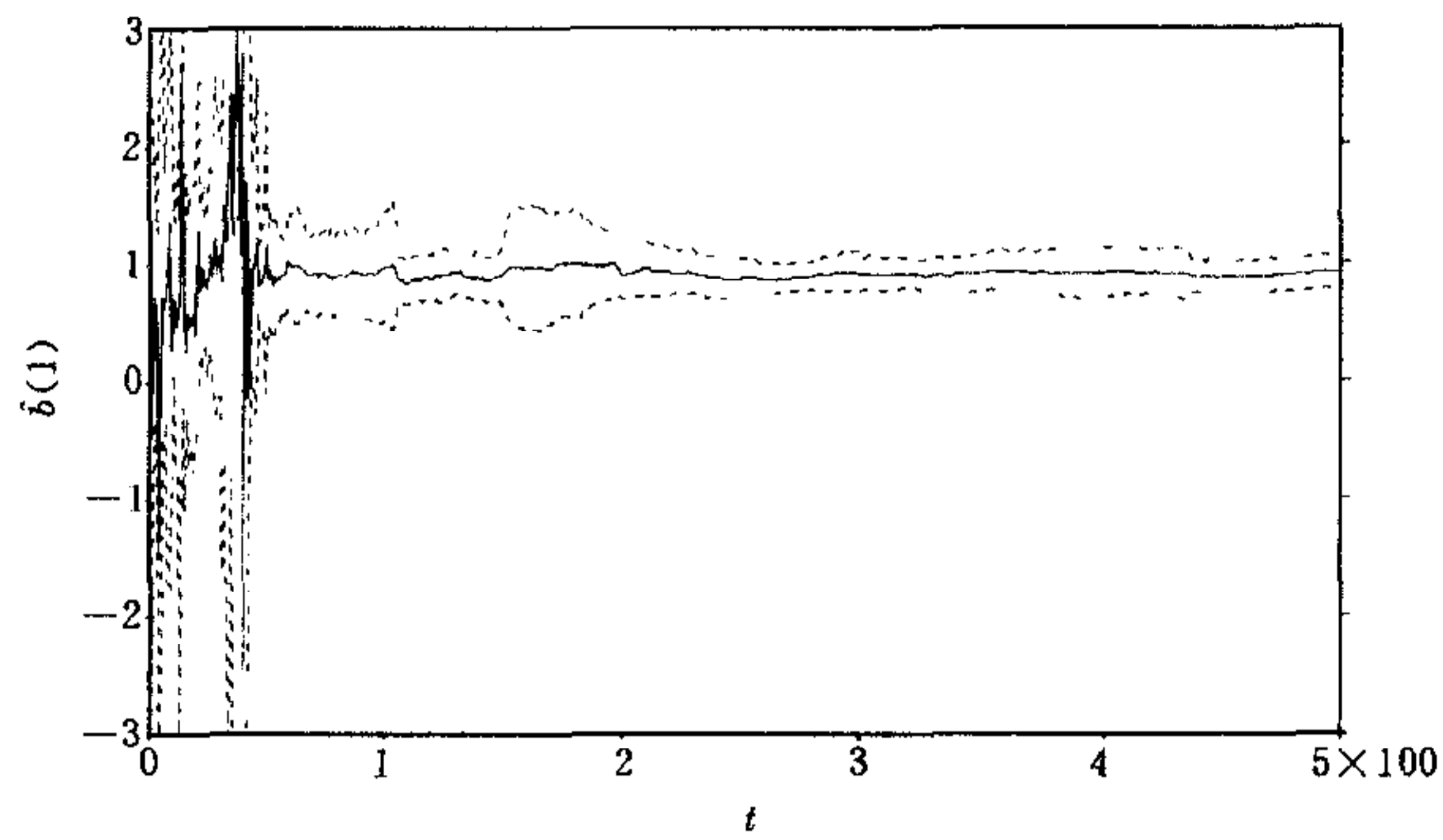
例 2. (时变系统的跟踪) 采用一个线性时变 MA 模型. 初始 MA 模型是

$$x(n) = w(n) - 1.5w(n-1) + 1.3w(n-2) + v(n).$$

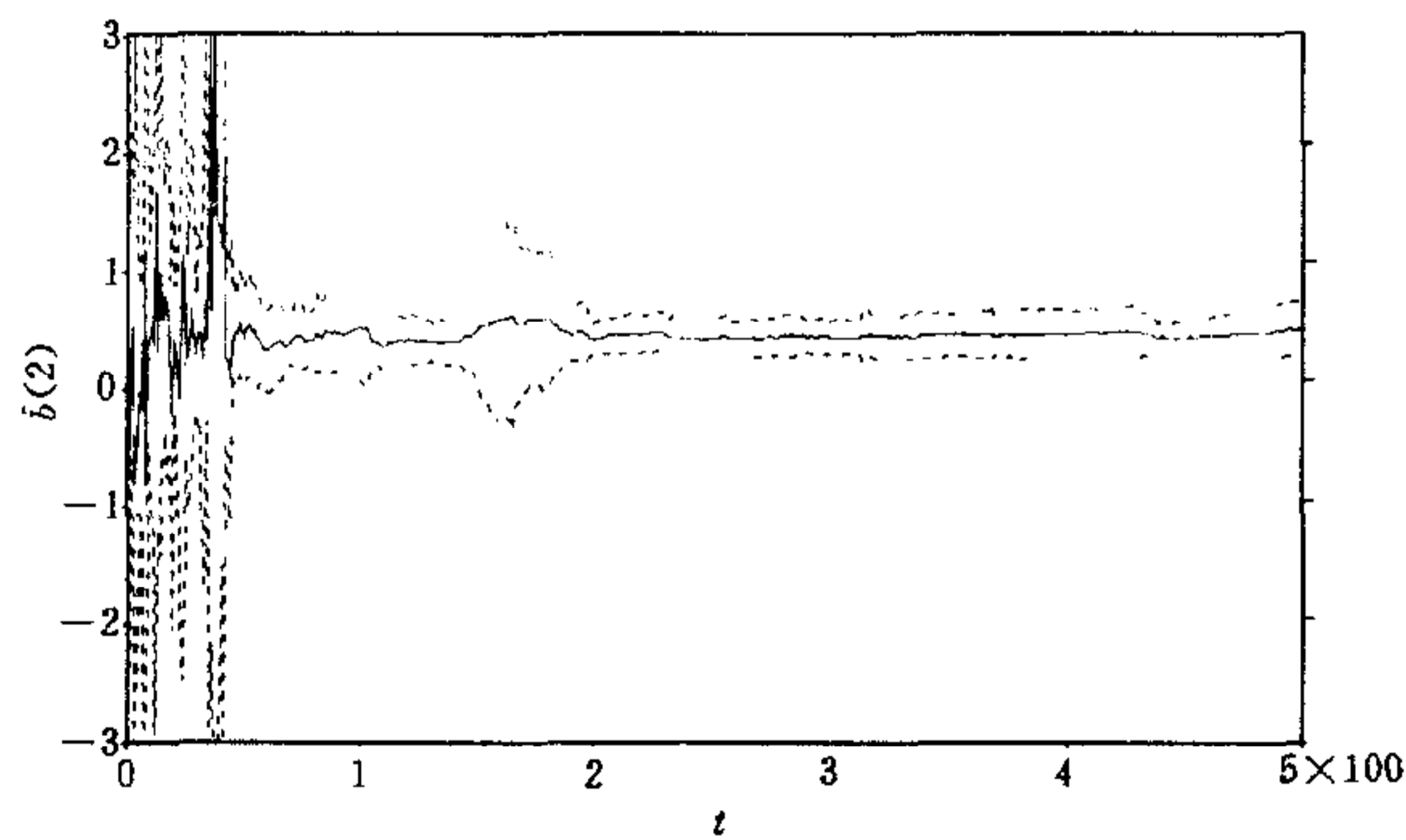
它的零点位于 $-0.75 \pm j0.86$. 这是一个最大相位 FIR 系统. 经过 5000 个数据点后 MA 模型变为

$$x(n) = w(n) - w(n-1) + 0.6w(n-2) + v(n).$$

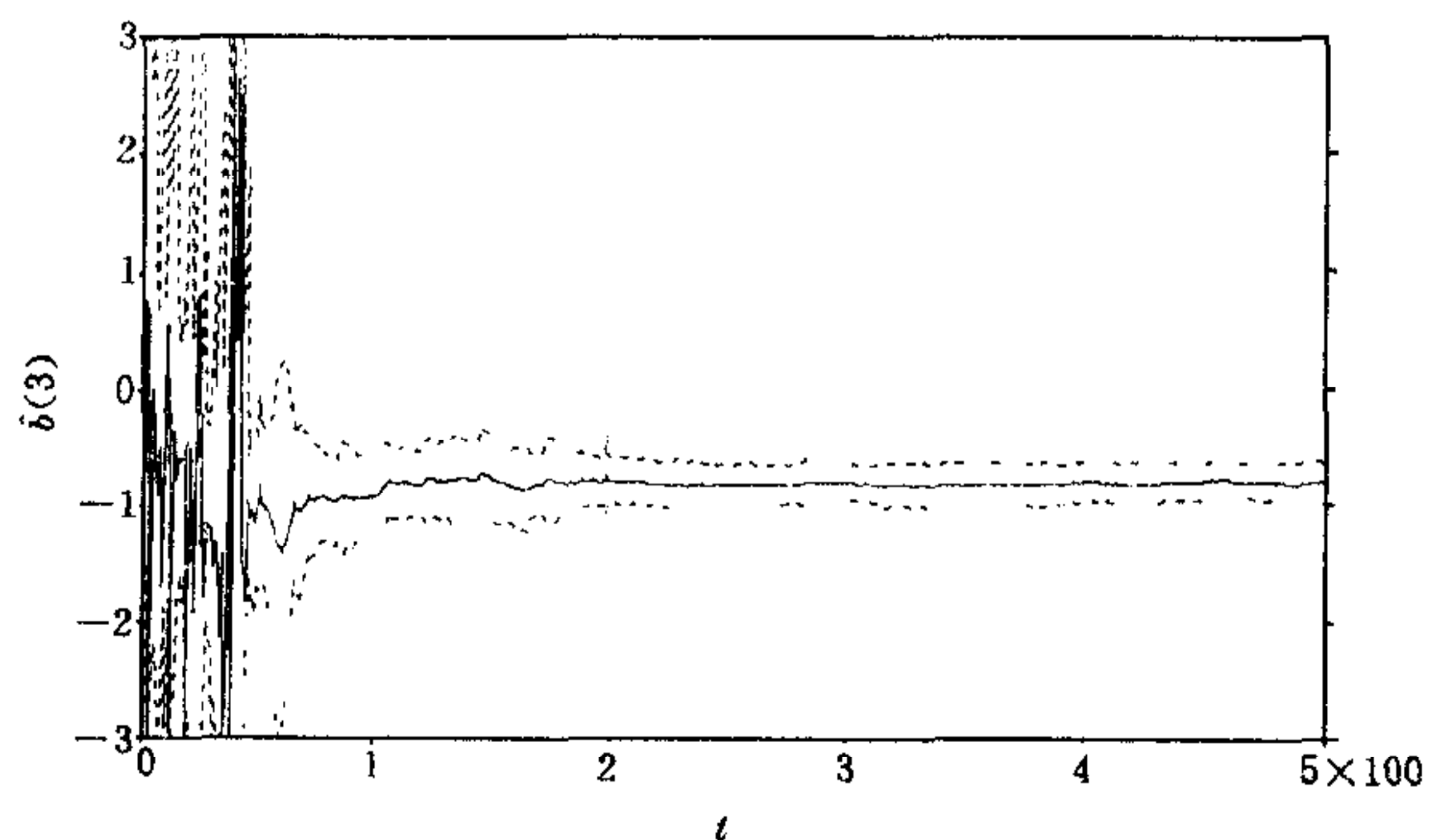
它的零点位于 $0.5 \pm j0.59$. 这是一个最小相位 FIR 系统. SNR=10 dB, 在三阶累量的递推更新中取式(9)中的 $\lambda=0.999$ 时的辨识结果如图2. 可以看出经过大约1500点的初始振荡后, 算法开始较好地跟踪系统参数的变化.



(a)



(b)

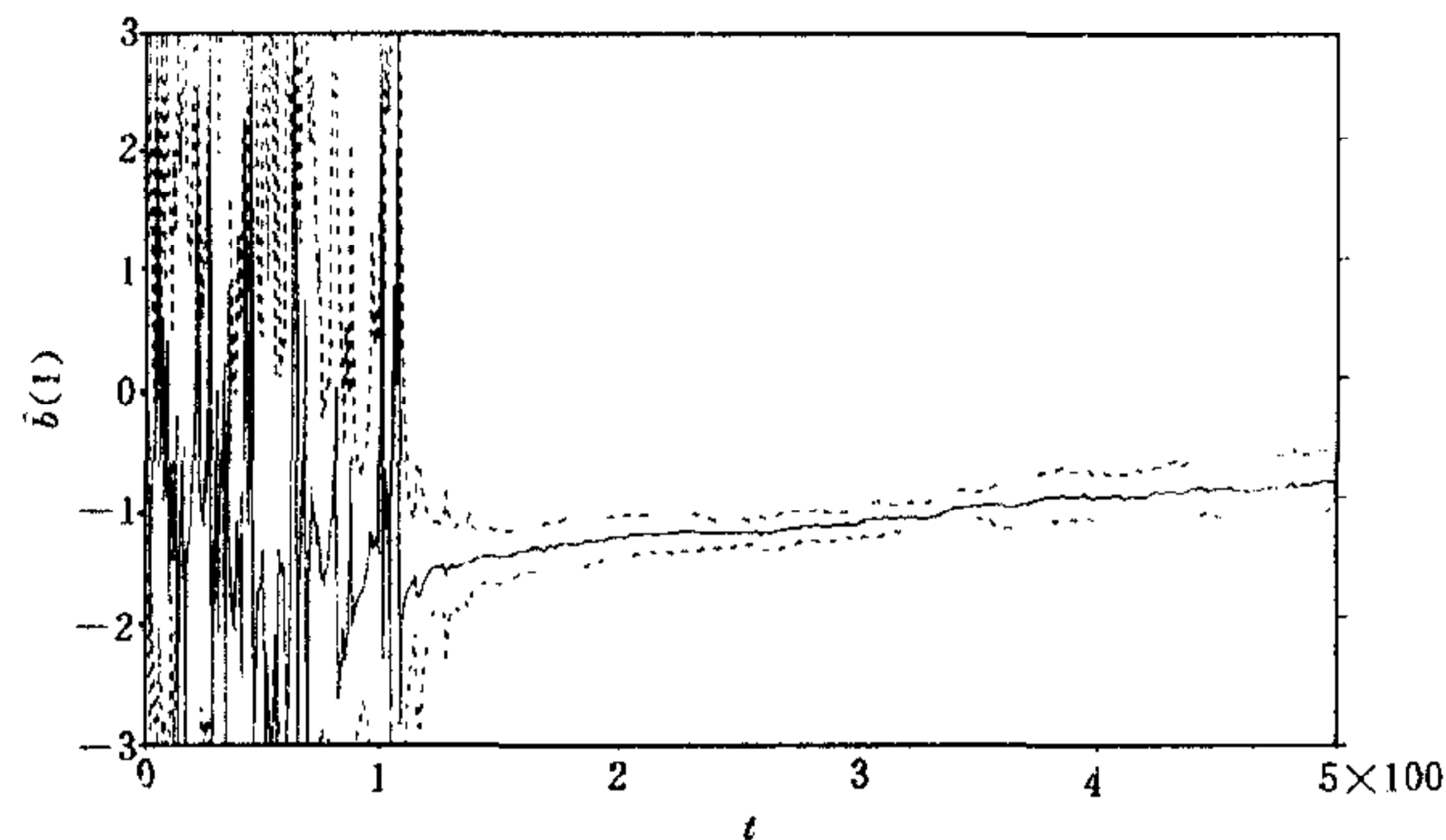


(c)

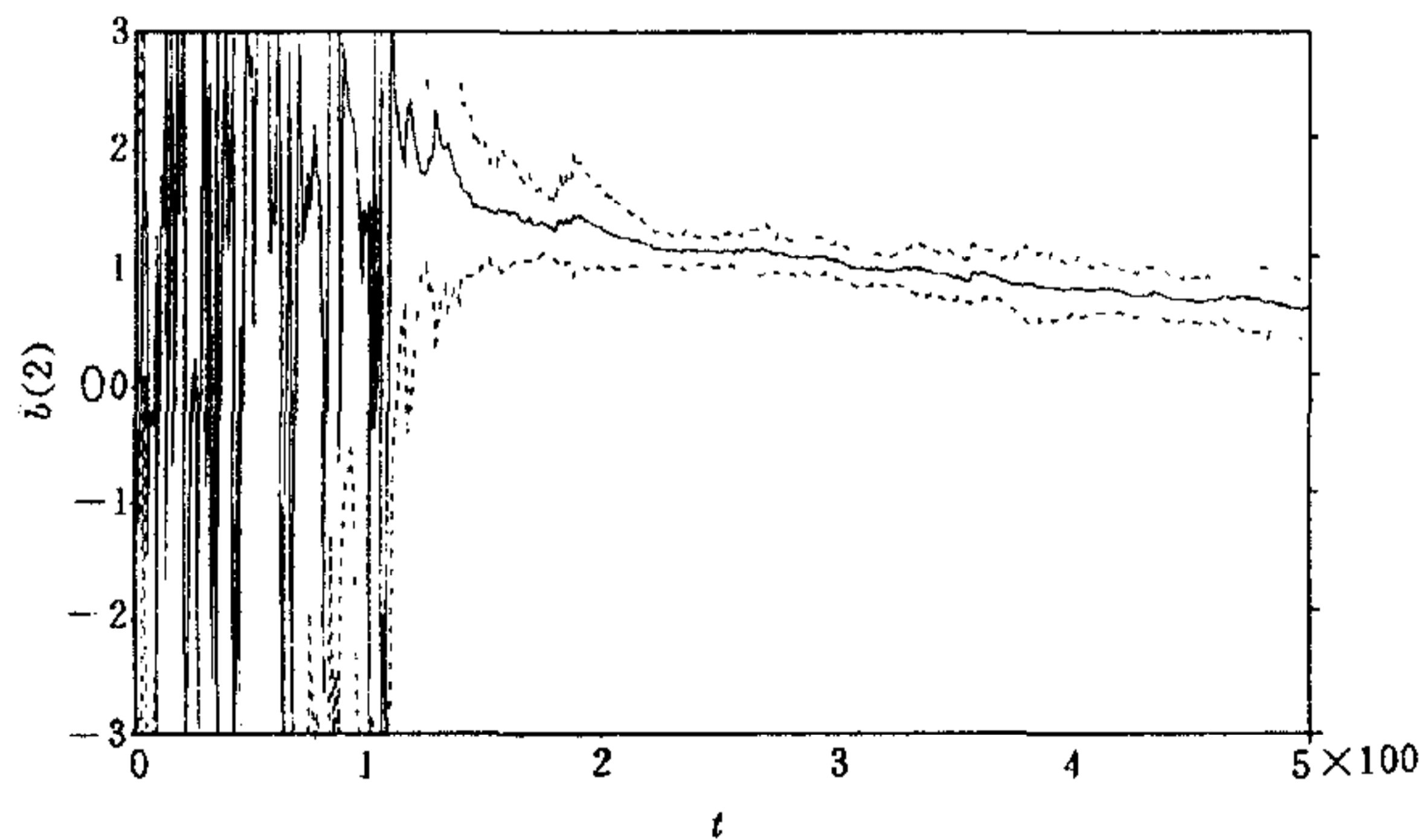
图1 自适应参数估计(均值±标准差)

SNR=5 dB

本文提出了一种基于线性代数方程组约束和梯度法的非最小相位 FIR 系统的自适应辨识算法. 由于只用高阶累量, 所以可以抑制高斯观测噪声的影响. 选用的线性方程组具有列满秩的系数矩阵, 从而保证了系统参数的唯一可辨识性. 本文还重点讨论了梯度法中步长的选择, 提出了收敛速度意义上的最佳变步长. 整个算法的计算量较小(包括步长的计算). 仿真实验结果表明了算法的有效性. 这种算法克服了 ORIV 算法中矩阵分解的困难, 可以适用于所有受线性代数方程组约束的系统辨识问题.



(a)



(b)

图2 时变参数的跟踪(均值±标准差)

SNR=10 dB, $\lambda=0.999$

参 考 文 献

- [1] Mendel J M. Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. In: *Proc. IEEE*, 1991, **79**(3): 278—305.
- [2] Zhang X D, Zhang Y S. FIR system identification using higher-order statistics alone. *IEEE Trans. Signal Processing*. 1994, **42**(10): 2854—2858.
- [3] Ljung L, Soderstrom T. Theory and practice of recursive identification. Cambridge, MA: The MIT Press, 1983.
- [4] Nikias C L, Mendel J M. Signal processing with higher-order spectra. *IEEE Trans. Signal Processing Mag.*, 1993, **10**(3): 10—37.
- [5] Friedlander B, Porat B. Adaptive IIR algorithms based on higher-order statistics. *IEEE Trans. Acoust., Speech, Signal Processing*, 1989, **37**(4): 485—495.
- [6] Chiang H H, Nikias C L. Adaptive decomposition and identification of nonminimum phase FIR systems based on

cumulants. *IEEE Trans. Automat. Contr.*, 1990, **35**(1): 36—47.

- [7] Haykin S. Adaptive Filter Theory, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [8] Luenberger D G. Linear, Nonlinear Programming. Reading, MA: Addison-Wesley, 1984, 214—220.
- [9] Scales L E. Introduction to Nonlinear Optimization. London: Macmillan, 1985, 19.
- [10] Golub G H, Van Loan C F. Matrix Computations. Baltimore, MD: The John Hopkins Univ. Press, 1983, 276—285.

ADAPTIVE IDENTIFICATION OF NONMINIMUM PHASE FIR SYSTEMS

SONG YU ZHANG XIANDA LI YANDA

(Dept. of Automation, Tsinghua University, Beijing 100084)

Abstract Based on linear equation constraints and steepest descent method, an adaptive algorithm is proposed for nonminimum phase FIR system identification. The uniqueness of identification is guaranteed by the full rank of the coefficient matrix of the utilized linear equations. Effects of colored Gaussian noise are automatically handled because of using merely higher-order cumulants. The optimal variable step sizes are discussed in the sense of convergence rate. Effectiveness of the new algorithm is verified by simulation results.

Key words Adaptive system identification, nonminimum phase system, higher-order cumulant.

宋 宇 1970年生于成都. 1989年于浙江大学电机系获学士学位, 1992年于电子科技大学电子工程系获硕士学位. 现为清华大学自动化系博士研究生. 主要研究兴趣是自适应信号处理及在数字通信中的应用.

张贤达 1969年毕业于原西安军事电讯工程学院, 1982年获哈尔滨工业大学工学硕士学位, 1987年获日本东北大学工学博士学位, 1990年至1991年在美国加州大学(San Diego)博士后一年. 现任清华大学自动化系教授, 博士生导师. IEEE 高级会员, 纽约科学院 Member, 近年在 IEEE 的 SP, AC 和 IT 三家汇刊发表论文十多篇, 出版专、译著三本. 主要研究领域为现代信号处理, 尤其是非高斯信号处理、循环平稳时间序列分析和智能信号处理.