

# 语音信号同伦非线性模型分析理论与算法<sup>1)</sup>

韦 岗 张丽清 李向武 欧阳景正

(华南理工大学无线电与自控研究所 广州 510641)

**摘 要** 提出了一种基于同伦理论的语音信号非线性模型分析理论与算法,具有计算量小、稳定性强、收敛性好等一般非线性模型分析方法所没有的优点,同时保持了线性模型的简单性与通用性.实验结果证实了该方法的有效性.

**关键词** 语音信号建模,非线性,同伦.

## 1 引言

语音信号建模是语音识别、合成及编码的重要基础.早期由于计算技术的限制,主要采用线性自回归(全极点)模型来描述语音信号<sup>[1]</sup>,并在此基础上提出了线性预测、同态解卷等一系列语音信号分析方法<sup>[2]</sup>.事实上,由于信息信号的产生是一个非线性过程<sup>[3]</sup>,其中还存在着混沌现象<sup>[4]</sup>,因而用线性模型来描述语音信号在理论上是不合适的.近年来,人们已越来越重视语音非线性分析方法的研究,如神经网络方法<sup>[5,6]</sup>.但是由于非线性模型分析方法普遍受到计算复杂、稳定性差等问题的困扰,目前还未得出一种令人接受的语音信号非线性模型分析方法.

语音信号的全极点模型虽不十分准确,但由于它具有计算简单及通用性好的优点,因而仍得到了广泛的应用.为了克服线性及非线性模型分析各自的不足,又保持其优点,本文在同伦理论的基础上,提出了一种同伦非线性模型分析方法.理论分析及实验结果均表明,它是一种非常有效的语音信号非线性模型分析方法.

## 2 语音信号模型分析

目前使用得最广泛的语音信号数学模型是全极点模型,其时域描述如下:

$$x(n) = \sum_{i=1}^p a_i x(n-i) + e(n). \quad (2.1)$$

其中  $x(n)$  为语音抽样信号,  $e(n)$  为激励信号,  $\{a_i\}$  为模型参数(又称线性预测系数),  $p$  为模型阶数.式(2.1)对应的频域形式为

$$X(Z) = \frac{E(Z)}{1 - \sum_{i=1}^p a_i Z^{-i}}. \quad (2.2)$$

1)国家自然科学基金、国家教委博士点基金及广东省自然科学基金资助课题.

上述线性模型分析要解决的问题是如何由语音信号来决定模型系数. 考虑到语音信号的时变性, 人们常采用短时分析技术, 即每一小段(通常是20—30 ms)内求解一个模型系数, 采用如下的最小均方误差解法.

定义短时平均预测误差为

$$E_m = \sum_n \left( x_m(n) - \sum_{i=1}^p a_i x_m(n-i) \right)^2, \quad (2.3)$$

其中  $x_m(n) = x(n+m)$  表示是在抽样点  $m$  附近选取的一个语音段. 在式(2.3)中令  $\partial E_m / \partial a_k = 0, k=1, 2, \dots, p$ , 就可以得到一组线性方程组, 解之即可得出最佳的模型参数.

由上述分析可见, 线性预测模型最大的优点是模型求解是一个线性问题, 容易计算. 缺点是模型精度不高, 不能描述语音信号的非线性性, 因而导致在此基础之上的语言处理效果不理想. 非线性模型虽有可能更精确, 但其求解是一个非线性寻优过程, 计算量大, 收敛性及稳定性均无法保证. 此外还不易得出一个通用的非线性模型.

为解决上述问题, 采用如下的非线性模型:

$$x(n) = f\left(\sum_{i=1}^p a_i x(n-i)\right) + e(n). \quad (2.4)$$

其中  $f$  为一个可随信号性质不同而改变的非线性函数. 显然, 式(2.1)所示的线性模型只是式(2.4)所示模型取  $f(y) = y$  时的特例.

式(2.4)所示的非线性模型决定于两个因素, 其一是非线性函数  $f$  的形式, 其二是模型参数  $\{a_i\}$  的选取. 为满足语音信号的时变性, 这两个因素都应每隔一小段就变化一次. 只有在  $f$  的形式确定之后, 方能对  $\{a_i\}$  进行求解, 而这一般又是一个非线性寻优过程.

### 3 同伦基本理论

同伦理论是解决非线性优化问题的新理论<sup>[7,8]</sup>, 同伦方法的基本思想是构造一个同伦函数, 将一个已知优化解的问题与待解的优化问题联系起来, 然后从已知优化解开始, 利用同伦参数的变化, 最终求得待解问题的最优解. 设  $f \in C^1(R^n, R^n), (n \geq 1)$  是一个非线性映射,  $f_0 \in C^1(R^n, R^n)$  有一个已知解  $t_0$ . 构造同伦函数  $H(t, \lambda) \in C^1(R^{n+1}, R^n)$ , 满足

$$H(t, 0) = f_0(t), \quad H(t, 1) = f(t), \quad t \in R^n, \quad (3.1)$$

其中  $f_0(t)$  的解已知,  $f(t)$  的解待求,  $\lambda$  为同伦参数. 一个点  $y = (t, \lambda) \in R^{n+1}$  称为  $H$  的正则点, 如果  $H$  在  $y$  点上的 Jacobi 矩阵的秩为  $n$ . 一个点  $z \in R^n$  称为  $H$  的正则值, 如果对于任一  $y \in H^{-1}(z)$  都是  $H$  的正则点.

若  $0$  是同伦函数  $H$  的正则值, 则  $H^{-1}(0) = \{(t, \lambda) | H(t, \lambda) = 0\}$  是一维流形. 因此只需要从  $(t_0, 0)$  开始跟踪同伦曲线  $H^{-1}(0)$  直到  $\lambda=1$  为止. 如果  $H^{-1}(0)$  中没有曲线连接  $(t_0, 0)$  到截面  $\lambda=1$ , 那么同伦方法失效. 所以同伦函数的构造是有效实现同伦方法的重要因素, 既要保证  $0$  是同伦函数的正则值, 又要求同伦曲线  $H^{-1}(0)$  能连接  $(t_0, 0)$  到截面  $\lambda=1$ . 通常同伦函数的构造方法是用线性连接, 如

$$H(t, \lambda, t_0) = \lambda f(t) + (1 - \lambda)(t_0 - t), \quad (3.2)$$

或

$$H(t, \lambda, t_0) = \lambda f(t) + (1 - \lambda)(f(t) - f(t_0)), \quad (3.3)$$

其中  $t_0$  是  $H$  在  $\lambda=0$  时的零点.

**定理1.** 设  $H(t, \lambda, t_0): R^n \times R \times R^n \rightarrow R^n$  是一个光滑映射, 如果  $0 \in R^n$  是  $H$  的正则值, 则对于几乎所有的  $t_0 \in R^n$ ,  $0$  为限制映射  $H_{t_0}(\cdot) = H(\cdot, \cdot, t_0): R^n \times R \rightarrow R^n$  的正则值.

根据该定理知, 对几乎所有的  $t_0 \in R^n$ ,  $0$  是限制映射  $H_{t_0}$  的正则值. 为了保证同伦曲线  $H^{-1}(0)$  与截面  $\lambda=1$  相交, 可以对同伦函数的构造加以限制. 设  $D$  是包含  $t_0$  的有界域, 1)  $t_0$  是  $H$  在  $D \times \{0\}$  上的唯一零点; 2)  $H(t, \lambda, t_0) \neq 0$ , 对  $(x, \lambda) \in \partial D \times [0, 1]$ . 这时由于对于几乎所有的  $t_0 \in R^n$ ,  $H^{-1}(0)$  是由正则点组成的光滑曲线, 另外,  $H^{-1}(0)$  又不准停留在  $\lambda < 1$  处, 则必然有  $H^{-1}(0)$  与截面  $\lambda=1$  相交.

由于同伦曲线  $H^{-1}(0)$  有可能出现转向点、分支点、高阶零点等奇性, 在实现时往往采用  $H^{-1}(0)$  的弧长  $s$  为参数, 进行跟踪计算. 关于  $H(t(s), \lambda(s))$  对弧长的微分, 有

$$A(s)\dot{W}(s) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.4)$$

其中

$$\dot{W}(s) = \frac{dW}{ds}, A(s) = \begin{bmatrix} H'(\omega(s)) \\ \dot{\omega}(s)^T \end{bmatrix} = \begin{bmatrix} H'_t(\omega(s)) & H'_\lambda(\omega(s)) \\ \dot{t}(s)^T & \lambda(s) \end{bmatrix}.$$

关于方程(3.4)的解, 有如下定理:

**定理2.** 若  $H \in C^1(R^{n+1}, R^n)$ ,  $0$  是  $H$  的正则值, 那么矩阵  $A(s)$  在  $H^{-1}(0)$  上是非奇的.

**证明.** 反证法. 设对于某  $s_0$ ,  $A(s_0)$  奇异, 则存在  $(\xi^T, \tau) \neq 0$ , 使得

$$A(s) \begin{bmatrix} \xi \\ \tau \end{bmatrix} = \begin{bmatrix} H'(\omega(s)) \\ \dot{\omega}(s)^T \end{bmatrix} \begin{bmatrix} \xi \\ \tau \end{bmatrix} = 0, \quad (3.5)$$

由于  $0$  是  $H$  的正则值,  $H'(\omega(s))$  的零空间是一维的, 所以  $\dot{\omega}(s)$  是  $H'(\omega(s))$  零空间的一个基. 由(3.5)式知  $(\xi^T, \tau)$  是  $H'(\omega(s))$  零空间的向量, 所以存在  $q \neq 0$ , 使得

$$(\xi^T, \tau) = q\dot{W}(s).$$

代入(3.5)式第二式产生矛盾. 因此  $A(S)$  是非奇的.

同伦方法可以通过求解(3.4)式的微分方程初值问题  $\omega(0) = (t_0, 0)$  来实现. 具体实现时可以采用一种预估校正算法跟踪同伦曲线  $H^{-1}(0)$ , 它从初值  $\omega_0 = \omega(0) = (t_0, 0)$  开始, 计算曲线上一串点  $\omega_1, \omega_2, \dots$ , 使得每一点  $\omega_{i+1}$  是由  $\omega_i \in R^{n+1}$  进行预估和校正得出的. 预估估值可以用下述方法实现.

$$v_i = \omega_i + \mu_i u_i, \quad (3.6)$$

其中  $\mu_i > 0$  是一个预测步长,  $u_i \in R^{n+1}$  是方程

$$\begin{bmatrix} H'(\omega(S)) \\ (\mu_i)^T \end{bmatrix} u_i = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (3.7)$$

的解, 它满足  $(u_i)^T u_i > 0$ . 而校正算法可以选用 Newton 方法, 具体算法是解方程

$$\begin{cases} H'(v_k)\delta_k = -H(v_k), \\ (u_i)^T \delta_k = 0, \quad k = 0, \dots, \\ v_{k+1} = v_k + \delta_k. \end{cases} \quad (3.8)$$

这里  $v_0$  是由 (3.6), (3.7) 式得出的预估点, 从  $k=0$  开始迭代直到满足  $\|H(v_k)\| \leq \varepsilon$   $\|H'(v_0)\|$  时停止, 并令  $\omega_{i+1} = v_{k+1}$ .

同伦理论的进一步知识可参见文献[7,8].

## 4 语音信号的同伦非线性模型

根据上述同伦理论, 在式(2.4)中, 取  $f$  具有如下形式便得到了同伦非线性模型:

$$f(y) = (1 - \lambda)y + \lambda g(y). \quad (4.1)$$

其中  $0 \leq \lambda \leq 1$  为同伦参数,  $g$  为一固定的非线性函数.

式(4.1)揭示了同伦非线性模型之间的内在联系: 当  $\lambda=0$  时, 同伦非线性模型即为一般的线性模型, 当  $\lambda$  由 0 变到 1 时, 模型也随之由线性变成非线性, 且非线性程度不断增强. 将式(4.1)代入式(2.4), 可得出同伦非线性模型的时域表达式如下:

$$x(n) = (1 - \lambda) \sum_{i=1}^p a_i x(n-i) + \lambda g\left(\sum_{i=1}^p a_i x(n-i)\right) + e(n). \quad (4.2)$$

下面介绍同伦非线性模型的高效解法.

定义模型误差为

$$J = \sum_n e^2(n) = \sum_n \left( x(n) - f\left(\sum_{i=1}^p a_i x(n-i)\right) \right)^2. \quad (4.3)$$

当  $f$  确定后, 最佳的模型参数  $\{a_i\}$  可通过令  $\partial J / \partial a_k = 0, k=1, 2, \dots, p$  进行求解, 对于一般的非线性模型, 这是一个很困难的非线性寻优问题, 存在着初始值及寻优解法的选择问题. 对于同伦非线性模型, 其求解的基本方法也与上述方法一致. 但由于同伦模型在结构上的特点, 可采用如下步骤有效地完成模型的求解, 包括最优非线性函数  $f$  的确定及对应的最佳模型参数  $\{a_i\}$  的求解.

步骤1. 令  $\lambda = \lambda_1 = 0$ , 此时式(4.2)事实上是一个线性模型, 用传统的最小均方误差法求出最佳模型参数  $\{a_i\}$ , 并求出对应的模型误差.

步骤2. 令  $\lambda = \lambda_2 = \lambda_1 + \Delta\lambda$ , 以  $\lambda = \lambda_1$  时求出的模型参数  $\{a_i\}$  作为初始值, 用迭代法求解新的最佳模型参数  $\{a_i\}$ , 并求出对应的模型误差.

步骤3. 令  $\lambda_1 = \lambda_2$ , 重复以上步骤, 直至  $\lambda = 1$ .

步骤4. 对应于模型误差最小的  $\lambda$  及  $\{a_i\}$  即为最终的模型参数.

易见, 由于上述算法是从线性模型开始的, 因而不存在一般非线性模型求解中遇到的初值问题. 此外, 当每次  $\lambda$  值改变不大时, 其对应的最佳模型参数  $\{a_i\}$  也变化不大, 因此可用如 Newton 法、共轭梯度法等方法高效地完成步骤2, 直至完成整个模型求解过程.

## 5 实验结果

采用上述的非线性模型, 对中国科学院声学所制作的标准普通话语音数据进行了建模实验. 语音信号的取样率为 8 kHz, 每 20 ms 为一帧. 模型求解时,  $\Delta\lambda$  取为 1/16, 式(4.2)中非线性函数  $g$  选用如下两种:

$$g_1(y) = \frac{1 - e^{-y}}{1 + e^{-y}} \quad (5.1)$$

及

$$g_2(y) = y e^{(1-y^2)/2}. \tag{5.2}$$

其波形如图1所示. 注意  $g_1$  事实上即为神经网络中常用的 Sigmoid 函数.

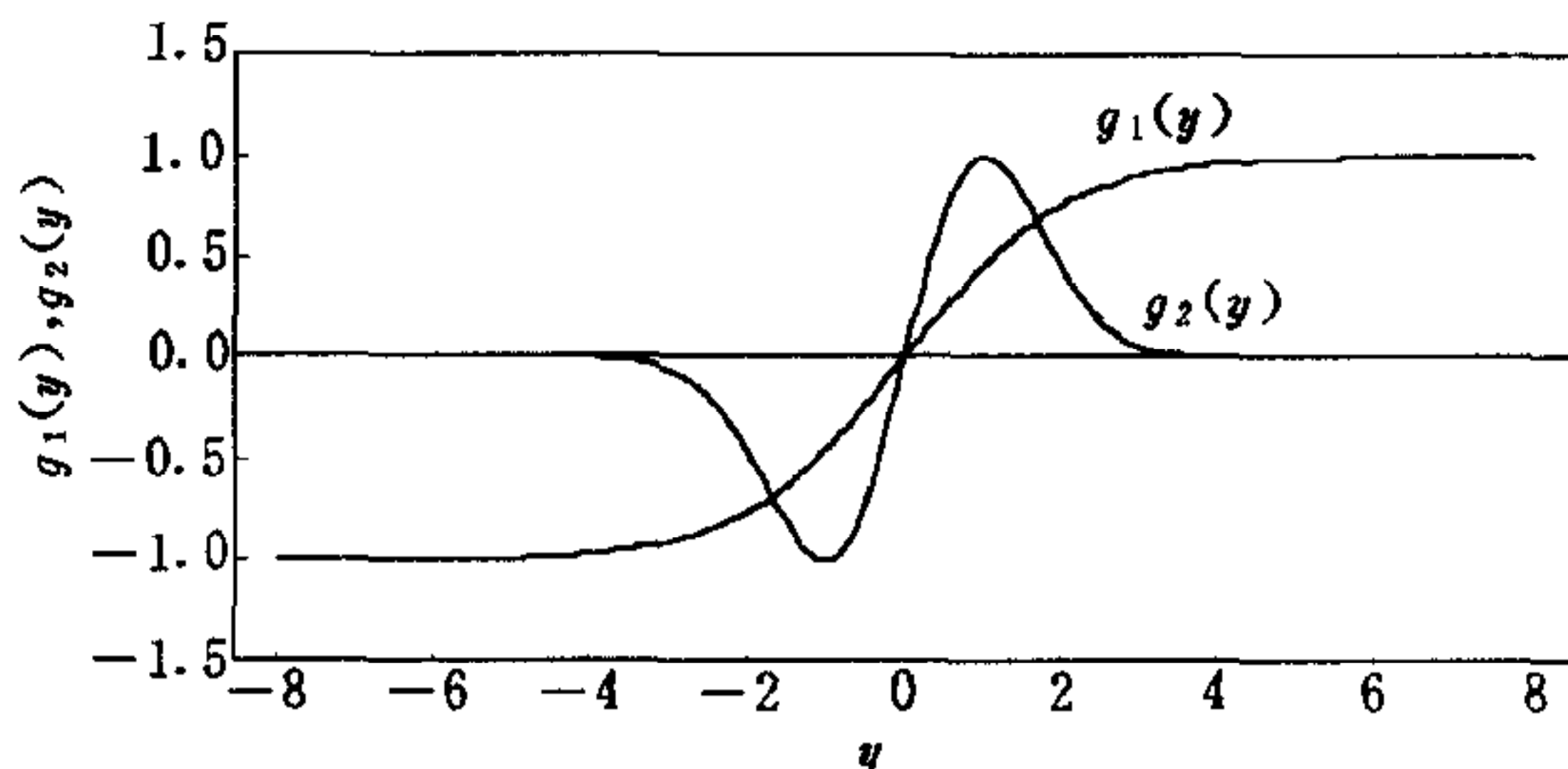


图1 非线性函数  $g_1(y)$ 、 $g_2(y)$  的波形图

为了便于比较,采用传统线性预测分析常用的模型阶数  $p=10,12$ ,采用如下归一化模型误差(即模型误差能量与信号能量之比)

$$E_0 = \sum_n e^2(n) / \sum_n x^2(n). \tag{4.3}$$

作为衡量指标,实验结果如表1所示. 图2为典型的同伦非线性模型的误差随  $\lambda$  值的变化曲线.

表1 语音信号同伦非线性建模实验结果(归一化误差  $E_0$ )

	线性模型	同伦非线性模型( $g=g_1$ )	同伦非线性模型( $g=g_2$ )
$p=10$ 模型误差均值	0.33273	0.23736	0.23708
$p=10$ 模型误差方差	0.21765	0.14931	0.14843
$p=12$ 模型误差均值	0.32897	0.23112	0.23103
$p=12$ 模型误差方差	0.21392	0.14526	0.14419

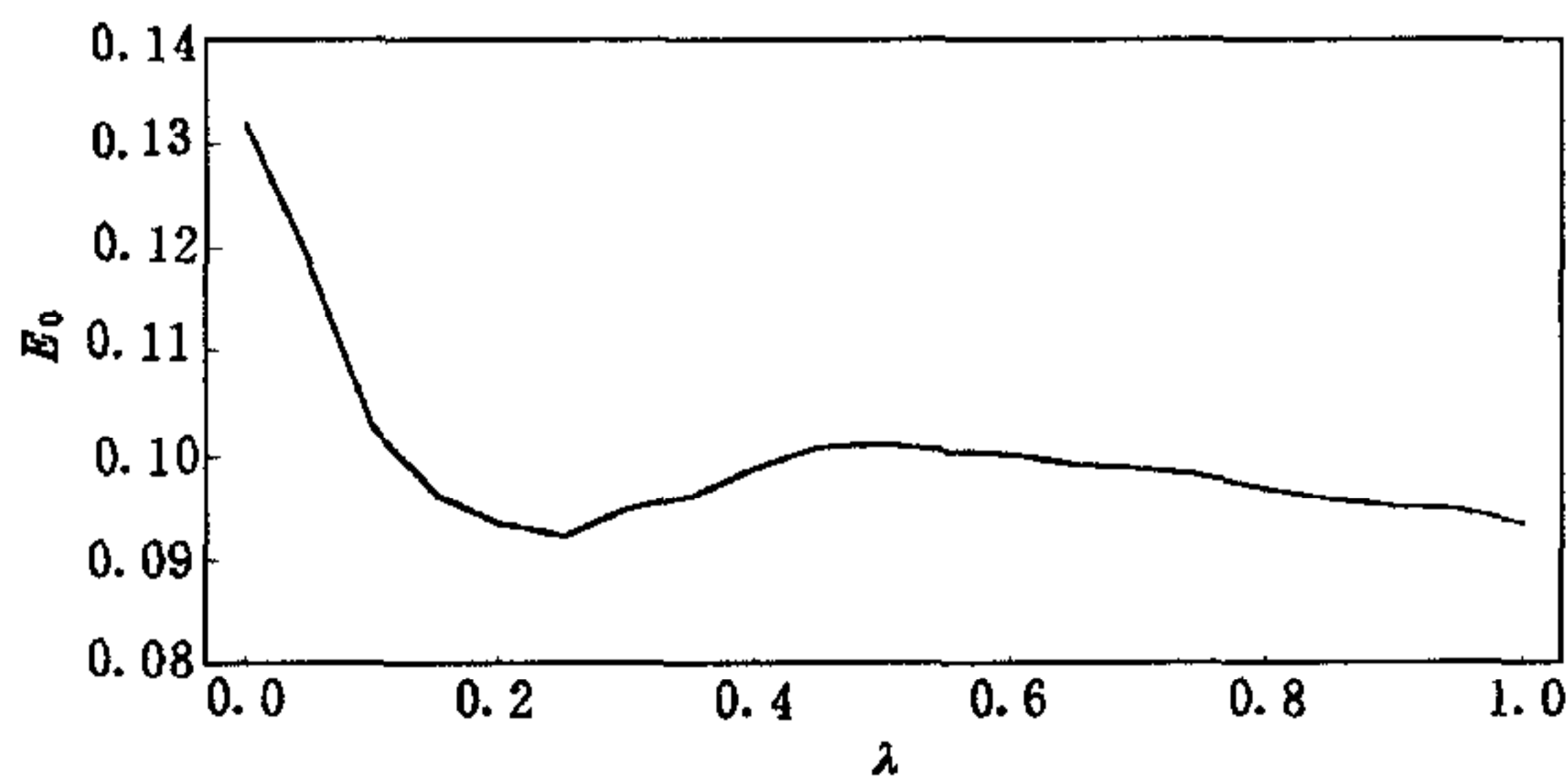


图2 同伦非线性模型的误差随同伦参数  $\lambda$  的典型变化曲线

由表1可见,当模型阶数相同时,同伦非线性的误差可减至线性模型的70%,并且不存在困扰一般非线性模型分析的计算问题,与传统的线性模型分析(相当于  $\lambda=0$  时的同伦非线性模型)有很好的兼容性,因而是一种理想的非线性模型. 此外,由于式(4.2)中的非线性函数  $g$  有无穷多种选取方法,因此本文方法的应用范围非常广泛.

## 参 考 文 献

- [1] Rabiner L, Juang B H. Fundamentals of Speech Recognition. New York: Prentice-Hall Int. Ltd., 1993.
- [2] Rabiner L, Schafer R W. 语音信号数字处理, 朱雪龙等译. 北京: 科学出版社, 1983.
- [3] Thompson C, Mulpur A and Mehta V. Transition to chaos in acoustically driven flow (acoustic streaming). *J. Acoust. Soc. Am.*, 1991, **90**:2097—2103.
- [4] Maragos P. Fractal aspects of speech signals; dimension and interpolation. In: Proc. IEEE ICASSP. 1991, 417—420.
- [5] 韦岗, 贺前华. 神经网络模型学习及应用, 北京: 电子工业出版社, 1994.
- [6] Kwong Sam, Wei Gang, Chan Yiukieng, Ouyang Jingzheng. Discrete utterance recognition based on nonlinear model identification with single-layer neural networks. In: Proc. IEEE Int. Symp. Circuits and Systems. 1993, 2019—2022.
- [7] Kalaba R, Tesfatsion L. Solving nonlinear equations by adaptive homotopy continuation. *Appl. Math. Comput.* 1991, **41**:99—115.
- [8] Zhang Liqing, Han Guoqiang. Optimal homotopy methods for solving nonlinear systems. *Numer. Math* 1993, **65**:523—538.

## HOMOTOPY NONLINEAR MODELING FOR SPEECH SIGNALS

WEI GANG ZHANG LIQING LI XIANGWU OUYANG JINGZHENG

*(Institute of Radio Engineering and Control, South China University of Technology, Guangzhou 510641)*

**Abstract** Based on the homotopy theory, this paper presents a novel nonlinear modeling method for speech signals, the homotopy nonlinear modeling method (HNMM). Unlike the ordinary nonlinear modeling methods, the HNMM is robust and easy to compute, and has very good convergence property. Besides, the HNMM is nearly as simple as the traditional linear modeling methods and thus can be used very extensively. The validity of the HNMM has been proved by experimental results.

**Key words** Speech signal modeling, nonlinearity, homotopy theory.

**韦岗** 1963年生. 1984年于清华大学无线电系获学士学位, 1987年、1990年于华南理工大学无线电与自动控制研究所分获硕士、博士学位, 毕业后留所任教. 现为教授、博士生导师、无线电与自动控制所副所长、电子与通信工程系副主任. 研究领域包括信号处理、模式识别、非线性系统理论及应用、神经网络等. 已出版学术著作四部, 发表论文40多篇.

**张丽清** 1961年生. 1983年于杭州大学数学系获学士学位, 1988年于中山大学计算机科学系获博士学位. 现为华南理工大学自动化系教授、博士生导师、系统工程研究所副所长. 研究领域包括神经网络、智能控制、非线性系统建模与仿真等. 发表论文40多篇.