

文章编号: 1005-8893(2000)03-0008-04

LMS 算法的最佳计算步长研究^{*}

王正洪

(江苏石油化工学院 计算机科学与工程系, 江苏 常州 213016)

摘要: LMS 算法在线性滤波中得到广泛应用。人们对其进行了许多研究, 给出了算法收敛的充分条件和对于固定计算步长的误差上下限算法。但以往对计算步长的定量研究尚很不充分。由于 LMS 算法中权向量输入数据的统计特性, 迭代计算中应突出使用较近的历史数据。另外, 由于 LMS 算法是噪声梯度法, 须使其计算步长逐渐缩小, 以保证 LMS 算法的稳态均方差趋向于系统的最小均方差。据此, 提出了一种估计最佳计算步长的新方法。新方法中将 LMS 算法中的各步权增量向量旋转后放至同一平面, 在此基础上利用凸函数外推法求出计算步长的当前最佳估计值。这样, 能使 LMS 算法的偏差较快地趋向于零。该法被用于汽车语言信号噪声消抵的计算机模拟, 结果令人满意。

关键词: LMS 算法; 偏差; 最佳计算步长; 最陡下降法

中图分类号: TN 911.72 TN 912.3

文献标识码: A

LMS 算法由于其简洁性, 在各种线性滤波器的设计中得到了广泛应用。由于 LMS 是随机梯度法, 故其稳态均方差比最小均方差大^[1]。因此, 在选取计算步长的同时, 不仅要考虑到使算法收敛, 还要考虑到使计算步长逐渐缩小^[2], 以保证 LMS 算法的稳态均方差趋向于系统的最小均方差。人们对 LMS 算法进行了许多研究, 给出了算法收敛的充分条件和对于固定计算步长的误差上下限算式^[3,4]。但以往的方法没有能充分利用计算历史数据选取计算步长。LMS 算法是由求凸函数最小值的最速下降法而导出的^[5], 因此, 为了合理地选取其计算步长, 必须对最速下降法作进一步的分析。

1 改进的最陡下降法

设 $z=f(\mathbf{x}) \equiv f(x_1, x_2, \dots, x_n)$ 为凸函数, 且已知其最小值点位于某一有限区域中。为求最小值, 可用最陡下降法。具体方法是:

P1 取定一初始点 $\mathbf{x}^{(0)}$, 并令 $p=1$;

P2 计算函数 $f(\mathbf{x})$ 在点 $\mathbf{x}^{(p)}$ 处的梯度

$$\nabla f(\mathbf{x}^{(p)}) = \left[\frac{\partial f(\mathbf{x}^{(p)})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x}^{(p)})}{\partial x_n} \right]^T \quad (1)$$

P3 判断是否有

$$\|\nabla f(\mathbf{x}^{(p)})\| < \epsilon \quad (2)$$

其中 ϵ 为预定的小正数。如果上式成立, 则计算结束, 否则继续进行如下的计算步骤;

P4 令 $\mathbf{x} = \mathbf{x}^{(p)} - t \nabla f(\mathbf{x}^{(p)})$, $t > 0$, 转 P5;

P5 选取一个计算步长 μ , 保证有 $f[\mathbf{x}^{(p)} - \mu \nabla f(\mathbf{x}^{(p)})] < f(\mathbf{x}^{(p)})$ (3)

令

$$\mathbf{x}^{(p+1)} = \mathbf{x}^{(p)} - \mu \nabla f(\mathbf{x}^{(p)}) \quad (4)$$

P6 用上式计算出 $\mathbf{x}^{(p+1)}$ 之后, 将 $\mathbf{x}^{(p+1)}$ 赋给 $\mathbf{x}^{(p)}$, 将 $p+1$ 之值赋给 p , 转 P2。

关于 μ 的取值方法, 可令

$$\mu \leftarrow \mu\varphi, \quad \varphi > 1, \quad \text{如 } f(\mathbf{x}^{(p+1)}) < f(\mathbf{x}^{(p)})$$

$$\mu \leftarrow \mu\beta, \quad 0 < \beta < 1, \quad \text{如 } f(\mathbf{x}^{(p+1)}) \geq f(\mathbf{x}^{(p)})$$

以使 (3) 式成立, 重新进行本次的迭代计算工作。

但 φ 和 β 究竟取多大合适值得研究。以往人们只是给出了它们的粗略值, 本文提出如下改进方法。

* 收稿日期: 2000-08-15

作者简介: 王正洪 (1944-), 男, 上海人, 副教授, 主要从事电工电子方面的研究。

考察已完成的所有迭代的计算情况。设第 i 次迭代所得的 x 的增量为 $\Delta x^{(i)}$ ，即有

$$x^{(i)} = x^{(i-1)} + \Delta x^{(i)}, i = 1, 2, \dots \tag{5}$$

设当前已完成了 p 次迭代。在平面直角坐标系 oxy 中，作点列

$$(x^{(i)}, y^{(i)}), i = 0, 1, 2, \dots, p \tag{6}$$

使有

$$(1) \quad y^{(i)} = f(x^{(i)}) \tag{7}$$

$$(2) \quad \sqrt{(x^{(i)} - x^{(i-1)})^2 + (y^{(i)} - y^{(i-1)})^2} = \|\Delta x^{(i)}\| \tag{8}$$

在 oxy 平面上，取一初始点 $(x^{(0)}, y^{(0)})$ ，只要有 $y^{(0)} = f(x^{(0)})$ 即可。然后，当完成了第 i 步迭代后，以点 $(x^{(i-1)}, y^{(i-1)})$ 为圆心，以 $\|\Delta x^{(i)}\|$ 为半径，画圆弧与水平线 $y = f(x^{(i)})$ 相交，其交点即为所要求的点 $(x^{(i)}, y^{(i)})$ 。这样就可作出 (6) 式所示的点列。

为确定第 p 次迭代中所用的计算步长 μ ，用一个抛物线来拟合点列 (6)，即找一个二次函数 $y = a_px^2 + b_px + c_p$ 来拟合点列 (6)。可用最小二乘法进行回归求得参数 a_p, b_p, c_p 。函数 y 的最小点和最小值分别是 $x_{min}^{(p)} = -\frac{b_p}{2a_p}$ 和 $y_{min}^{(p)} = -\frac{b_p^2}{4a_p} + c_p$ 。

令
$$x^{(p+1)} = x_{min}^{(p)}, y^{(p+1)} = y_{min}^{(p)} \tag{9}$$

由上式、(8) 式及 (4) 式可得

$$\mu \|\nabla f(x^{(p)})\| = \frac{\sqrt{(x_{min}^{(p)} - x^{(p)})^2 + (y_{min}^{(p)} - y^{(p)})^2}}{\|\nabla f(x^{(p)})\|} \tag{10}$$

由此得

$$\mu = \frac{\sqrt{(x_{min}^{(p)} - x^{(p)})^2 + (y_{min}^{(p)} - y^{(p)})^2}}{\|\nabla f(x^{(p)})\|} \tag{11}$$

上式所示即为第 p 次迭代中 μ 的最佳估计值。为了将本节方法用于 IMS 算法的改进，必须考虑 IMS 算法的特点。

2 改进的 IMS 算法

LMS 算法被用于求取线性滤波中自适应网络的加权因子。如图 1 所示为一个典型的横向自适应网络结构，其中 T 为延时器。第 n 个时刻该网络的输出为

$$d(n) = \sum_{i=0}^{N-1} w_i(n) \cdot x(n-i) \tag{12}$$

待辨识网络在第 n 个时刻的输出被假定具有形式 $d(n) = \sum_{i=0}^{N-1} h_i \cdot x(n-i)$ ，而

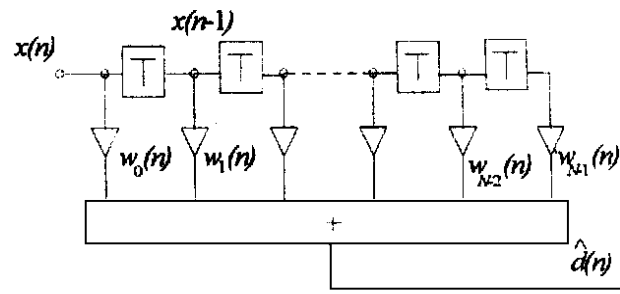


图1 横向自适应网络结构

$$e(n) = d(n) - d(n) \tag{13}$$

在 LMS 算法中，权向量的修正算式是

$$w(n+1) = w(n) + 2\mu e(n) x(n) \tag{14}$$

其中 μ 为计算步长或收敛速率。而

$$x(n) = [x(n-N+1), x(n-N+2), \dots, x(n)]^T \tag{15}$$

而 $2e(n) x(n) = -\frac{\partial e^2(n)}{\partial w(n)} - \hat{\nabla}_w [\epsilon(n)]$ 是 $-\nabla_w [\epsilon(n)] = -\frac{\partial E\{e^2(n)\}}{\partial w_N}$ 的估计值，因此 LMS 是最陡梯度法的一种近似。

仿照前述对 LMS 算法作出改进。必须注意对 LMS 算法来说有如下几点较为特殊：

(1) 为了确保 LMS 算法收敛，须有 $0 < \mu < \frac{1}{\lambda_{max}}$ ，这里 λ_{max} 是自相关矩阵 R 的最大特征值；

(2) 由于输入数据的统计特性，因此，当前迭代计算中应突出使用较近的历史数据；

(3) IMS 算法是噪声梯度法，故其稳态均方差比最小均方差大。因此，须使计算步长逐渐缩小，以保证 IMS 算法的稳态均方差趋向于系统的最小均方差；

(4) 还是因为 LMS 算法是噪声梯度法，在其计算过程中允许最小均方差的估计值有一定的上下波动。

综上所述，我们提出改进的 IMS 算法如下：

Q1 相应于某一初始时刻，取定初始权向量 $w_N(0)$ ，并令 $n=0$ ；

Q2 计算
$$\hat{\nabla}_w [\epsilon(n)] = \frac{\partial e^2(n)}{\partial w(n)} = -2e(n) x(n) \tag{16}$$

Q3 如果 $n \text{ Mod } M = 0$ （这里 M 为预先取定的正整数），则计算 $R = E\{x(l)x^T(l)\}$ 的估计

矩阵 R , 并求得其最大特征值的近似值 λ_{\max} ;

Q4 当 $n < n_0$ 时, 直接取 μ 为某一小于 $0.5\lambda_{\max}$ 的正数而不去考虑权向量增量的历史数据。这里, n_0 为某正整数, 其量级与 N 相当。不妨取 $n_0 = \max(N, 5)$;

Q5 当 $n > n_0$ 时, 建立直角坐标系 oxy , 在其中作点列

$$(x^{(i)}, y^{(i)}), i=1, 2, \dots, n-1 \quad (17)$$

使有

$$(1) \quad y^{(i)} = E \{e^2(l)\} = \frac{1}{K} \sum_{j=0}^{K-1} e^2(i-j) \quad (18)$$

其中, $i=1, 2, \dots, n-1$; K 为某正整数。

$$(2) \quad \sqrt{(x^{(i)} - x^{(i-1)})^2 + (y^{(i)} - y^{(i-1)})^2} = \|\Delta w^{(i)}_N\| \quad (19)$$

为强调较近迭代结果对 μ 的影响, 用一个二次函数 $y = a_n x^2 + b_n x + c_n$ 来拟合点列 (17) 中最后 n_1 个点所构成的子点列。这里, n_1 取为 5 与 10 之间的一个正整数。求出其最小值点和最小值分别为 $x_{\min}^{(n)}$

$$= -\frac{b_n}{2a_n} \text{ 和 } y_{\min}^{(n)} = -\frac{b_n^2}{4a_n} + c_n \text{ 令}$$

$$\mu_0^{(n)} = \frac{\sqrt{(x_{\min}^{(n)} - x^{(n-1)})^2 + (y_{\min}^{(n)} - y^{(n-1)})^2}}{\|\nabla E \{e^2(l)\}\|} \quad (20)$$

$$\mu_1^{(n)} = \min \left[\mu_0^{(n)}, \frac{0.5}{\lambda_{\max}}, \mu^{(n-1)} \right] \quad (21)$$

其中 $\mu^{(n-1)}$ 为上次迭代所采用的 μ 值。较为理想的情况是, 由于误差曲面的凸性及 $\nabla E \{e^2(l)\}$

对于 $\nabla E \{e^2(l)\}$ 的估计较好, 由 (20) 式所算的 $\mu_0^{(n)}$ 自动能逐次缩小, 从而当 n 较大时也会有

$$\mu_0^{(n)} < \frac{0.5}{\lambda_{\max}}, \text{ 从而有}$$

$$\mu_1^{(n)} < \mu_0^{(n)} \quad (22)$$

即这时按 (20) 式选取 μ 的第 n 次迭代值 $\mu^{(n)}$, 既能保证算法稳定收敛, 又能使 $\mu^{(n)} \rightarrow 0$ 。但是, 在一般情况下, (22) 式未必能成立。故可令

$$\mu^{(n)} = \begin{cases} \mu_1^{(n)}, & n \bmod M_0 > 0 \\ 0.95\mu^{(n-1)}, & n \bmod M_0 = 0 \end{cases} \quad (23)$$

这样既能保证算法稳定收敛, 又能使 $\mu^{(n)} \rightarrow 0$;

Q6 判断如下条件是否成立

$$\|w_N^{(n+1)} - w_N^{(n)}\| < \epsilon \text{ 或 } n > n_G \quad (24)$$

其中, ϵ 为预定的小正数; n_G 为预定的正整数, 用于限制总的迭代次数。一般, n_G 的范围可取为 1

$500 \leq n_G \leq 2500$, ϵ 的范围可取为 $10^{-6} \leq \epsilon \leq 10^{-5}$ 。

如果由 (24) 式所表示的条件成立, 则结束计算。否则, 进行赋值 $n \leftarrow n+1$, 再转 Q2。

3 分 析

参见 (12) 式至 (14) 式。实际输出与估计输出之间的均方差是

$$\zeta = E[d^2(l)] - 2w_N^T E[d(l)x] + w_N^T E[xx^T] w_N \quad (25)$$

将其最小值记为 ζ_{\min} , 将 IMS 算法在固定计算步长 μ 之下所得的稳态极限值记为 ζ_{ss} 。IMS 算法的偏差定义为

$$M_{IMS} = \frac{\zeta_{ss} - \zeta_{\min}}{\zeta_{\min}} \quad (26)$$

它可进一步表示为^[5]

$$M_{IMS} = \frac{\zeta_{ss} - \zeta_{\min}}{\zeta_{\min}} = \frac{S}{1-S} \quad (27)$$

其中

$$S = \sum_{i=1}^N \frac{\mu \lambda_i}{1 - \mu \lambda_i} \quad (28)$$

其中 $\lambda_i, i=1, 2, \dots, N$ 为自相关矩阵 R 的全体特征值。文献 [3] 中还给出了较为准确的 M_{IMS} 的上下限算式。不过由 (27) 式和 (28) 式已能看出, 当 $\mu \rightarrow 0$ 时, 且 LMS 算法迭代次数趋向于无穷大时, $M_{IMS} \rightarrow 0$ 。而上述改进的 IMS 算法能做到 $\mu \rightarrow 0$, 因而它能保证 $M_{IMS} \rightarrow 0$ 。由于此法不是简单地使用 $\mu \leftarrow \mu\beta$ ($0 < \beta < 1$) 来减小 μ , 所以可避免出现两个极端, 一是 μ 减小得太快使得收敛减慢 (如果 μ 保持为定值, 这意味着 ζ 接近 ζ_{ss} 所需的时间加长), 二是 μ 减小得太慢使得 M_{IMS} 的下限下降过慢。本文方法实际上是使 μ 根据以往的计算情况合理地自然减小。

4 算 例

对于在汽车中用电话与汽车外进行通讯时用自适应滤波器实现信号消抵的情况, 可用 21 权系数的低通滤波器来模拟汽车^[4]。这些系数可取为

$$\begin{aligned} h_1 &= 0.008697 = h_{21}, & h_7 &= 0.021363 = h_{15} \\ h_2 &= 0.026267 = h_{20}, & h_8 &= 0.015199 = h_{14} \\ h_3 &= -0.019490 = h_{19}, & h_9 &= -0.071330 = h_{13} \\ h_4 &= 0.007113 = h_{18}, & h_{10} &= 0.122376 = h_{12} \\ h_5 &= 0.013134 = h_{17}, & h_{11} &= 0.033457 = h_{11} \end{aligned}$$

$$h_6 = -0.028302 = h_{16} \quad (29)$$

使用 LMS 算法求含 M 个分量的权向量 w_N , 使得 $E\{e^2(n)\}$ 达最小。取 $M=N$ 。这里

$$e(n) = d(n) - \hat{d}(n) \quad (30)$$

$d(n)$ 如 (12) 式所示。由上式消抵后的信号即为语音信号。在使用改进的 LMS 算法时, 为减少计

算量, 运用 Honigt 等^[5]的研究结果, 将 $\frac{1}{\lambda_{\max}}$ 用

$\frac{2}{M \cdot \Phi_0} = \frac{2}{N \Phi_0}$ 替代。这里 $\Phi_0 = E\{x^2(n)\}$ 。实际计

算时用其估计值 $\hat{E}\{x^2(n)\}$ 。采用两种方法进行计算, 一为传统的 LMS 算法, 其中采用时变步长 μ , μ 的缩小简单地用 $\mu \leftarrow \mu\beta$ ($0 < \beta < 1$) 来实现。另一种方法则采用本文所述的改进的 LMS 算法。两种方法的学习曲线比较如图 2 所示。由图可见, 改进的 LMS 算法收敛速度较快。

5 结 论

本文所提出的改进的 LMS 算法, 充分运用了计算过程中的历史数据及误差曲面凸性的规律, 并对有关可能出现的特殊情况作了考虑 [参见 (22) 式、(23) 式], 因而能在确保收敛和不断减小偏差的前提下, 较充分地加快收敛速度。

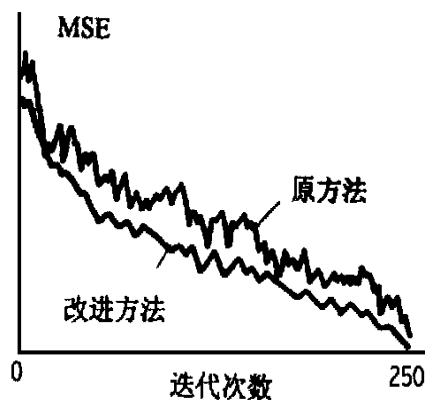


图2 学习曲线比较

参考文献:

- [1] 陈尚勤, 李晓峰. 快速自适应信息处理 [M]. 北京: 人民邮电出版社, 1993 2—47.
- [2] 王正洪. 语音信号噪声消抵的计算机模拟 [J]. 江苏石油化学学院学报, 1992, 4 (3, 4): 16—21.
- [3] Jaggi S, Martinez A B. Upper and Lower Bounds of the Misadjustment in the LMS Algorithm [J]. IEEE, 1990 (100): 164—166.
- [4] Horowitz L L. Performance Advantage for Complex LMS for Controlling Narrow Band Adaptive Arrays [J]. ASSP, 1981, 29 (6): 722—736.
- [5] Honig M J. Convergence Properties of an Adaptive Digital Lattice Filter [J]. ASSP, 1981, 29 (6): 642—653.

A Study on the Optimum Convergence Gain of LMS Algorithm

WANG Zheng—hong

(Department of Computer Science and Engineering, Jiangsu Institute of Petrochemical Technology, Changzhou 213016, China)

Abstract: LMS algorithm is used widely in various adaptive filters designing. A lot of work has been done on it. The adequate convergence condition of this algorithm has been got. And the upper and lower bounds of the misadjustment in the algorithm for a constant convergence gain have been got. The quantitative study on the convergence gain is yet insufficient. Because of the statistical characteristic of input data of the weights in a LMS algorithm, more attention should be paid to the recent historical calculation data. As the LMS algorithm is a kind of noise gradient method, the convergence gain should be lessened gradually to ensure that the quadratic mean deviation of the steady state takes the quadratic mean deviation of the system as its limit. According to this condition, a new method to estimate the optimum convergence gain is proposed. In the new method the increment of the weights vector in previous steps was rotated and placed to a plane, then the extended method for a one variable convex function was used. In this way, the misadjustment of the LMS algorithm can converge to zero quickly.

Key words: LMS algorithm; misadjustment; optimum convergence gain; steepest—descent method