

文章编号: 1672-2892(2010)05-0560-06

## 加权部分更新仿射投影算法

彭最亮, 李 锋

(复旦大学 电子工程系, 上海 200433)

**摘要:** 从数据重用因子出发, 得到了基于部分更新仿射投影算法(SR-APA)的改进算法。该算法通过加权改变了 SR-APA 的数据筛选规律, 从而降低了等效数据重用因子, 并且通过对未加权原始数据的重新利用巧妙地避免了加权带来的条件数增加问题, 最终达到了降低稳态均方误差(MSE)的效果。仿真结果表明, 该算法不仅 MSE 比 SR-APA 低, 收敛速度也比 SR-APA 快。在收敛速度相同时, 该算法计算量只有 SR-APA 计算量的 50% 左右。

**关键词:** 仿射投影算法; 部分更新; 加权; 条件数

**中图分类号:** TN911.72

**文献标识码:** A

## An affine projection algorithm with selective weighted regressors

PENG Zui-liang, LI Feng

(Department of Electronics Engineering, Fudan University, Shanghai 200433, China)

**Abstract:** Based on Affine Projection Algorithm with Selective Regressors(SR-APA), a new algorithm is proposed with a view to data reusing factor. The proposed algorithm introduces a weighting method to change the selective principles. Unweighted original data is reused in the iteration to avoid increasing the condition number of the input data, and finally the steady state Mean Square Error(MSE) is reduced. Simulation results show that the proposed algorithm not only reaches lower steady state MSE than SR-APA, but also accelerates the convergence speed of SR-APA. And with the same performance, the proposed algorithm has only 50% computations of SR-APA.

**Key words:** affine projection algorithm; selective regressors; weighted; condition number

仿射投影算法(APA)是一种重复利用过去数据以提高收敛速度的自适应滤波算法<sup>[1]</sup>。该算法在通信、自动控制等领域中有着广泛应用。在诸如动态噪声消除、回声消除等一些应用领域, 往往需要高阶的自适应滤波器来进行数据处理, 而高阶的 APA 的计算复杂程度与其输入数据的维数成正比, 高阶的 APA 会带来大量的计算, 因而会大大影响其实用性<sup>[2]</sup>。文献[3]提出的算法在每次权值更新时只更新滤波器的部分抽头, 在滤波器抽头数庞大时能减少计算量。集员滤波器的数据选择性也使得其计算量与 APA 相比大大降低<sup>[4]</sup>。部分更新仿射投影算法(SR-APA)从输入数据向量集合中选一个子集用于每次迭代, 该算法的最大优点在于所选择子集能够使代价方程的改变最小, 从而能够在大大降低传统 APA 计算量的同时, 最大程度保持传统 APA 的收敛性能<sup>[2]</sup>。然而 SR-APA 在改善稳态均方误差(MSE)方面却没什么效果。APA 的 MSE 与数据重用因子有关, 数据重用因子越高, 收敛越快, MSE 越大<sup>[5]</sup>。本文在 SR-APA 的基础上, 用加权的方法降低等效的数据重用因子, 同时保持传统 APA 的收敛特性, 降低 MSE, 提高 SR-APA 的性能。

### 1 仿射投影算法与部分更新仿射投影算法

#### 1.1 仿射投影算法

APA 的目标是为使下式最小化:

$$\begin{cases} \min \left\{ \frac{1}{2} \|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2 \right\}, \text{st.} \\ \mathbf{d}_k - \mathbf{X}_k \mathbf{w}_{k+1} = \bar{\mathbf{0}} \end{cases} \quad (1)$$

式中:  $\mathbf{w}_0 = [w_0(1), w_0(2), \dots, w_0(M)]^T$  为未知的  $M \times 1$  线性有限冲激响应(Finite Impulse Response, FIR)系统;  $\mathbf{x}_k = [x(k), x(k-1), \dots, x(k-M+1)]^T$  为  $M \times 1$  输入数据向量;  $\mathbf{X}_k = [\mathbf{x}_k, \mathbf{x}_{k-1}, \dots, \mathbf{x}_{k-L+1}]^T$  为  $L \times M$  输入数据矩阵,  $L$  为数据重用因子即 APA 的阶数;  $\mathbf{d}_k = [d(k), d(k-1), \dots, d(k-L+1)]^T$  为  $L \times 1$  期望输出,  $(\cdot)^T$  为矩阵的转置操作。

通过将约束最优化问题转化成无约束最优化问题, 可解得 APA 权值更新表达式为(详细证明见文献[1]):

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \mathbf{X}_k^T (\mathbf{X}_k \mathbf{X}_k^T)^{-1} \mathbf{e}_k \quad (2)$$

设  $\mathbf{y}_k = \mathbf{X}_k \mathbf{w}_k$  为实际输出,  $\mathbf{v}_k = [v(k), v(k-1), \dots, v(k-L+1)]^T$  为误差向量, 则上式中的误差向量  $\mathbf{e}_k = \bar{\mathbf{d}}_k + \bar{\mathbf{v}}_k - \bar{\mathbf{y}}_k$ ,  $\mu$  为步长因子。

此时的最优目标函数表达式为<sup>[2]</sup>:

$$J_{\min}(\mathbf{w}_{k+1}) = \frac{1}{2} \|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2 + \mathbf{e}_k^T (\mathbf{X}_k \mathbf{X}_k^T)^{-1} \mathbf{e}_k \quad (3)$$

### 1.2 部分更新仿射投影算法

设  $r \in (0, 1)$  为 1 个比例因子,  $R=rL$  为实际用于权值更新的输入数据向量维数,  $T_r = \{t_1, t_2, \dots, t_r\}$  为输入数据的  $R$  子集,  $t_1 < t_2 < \dots < t_r \in \{0, 1, \dots, L-1\}$ 。若所有  $T_r$  构成一个集合  $S$ , SR-APA 的目的即从  $S$  中选出一个最优的  $R$  子集  $T_{Ropt}$  用于权值更新<sup>[2]</sup>。改写约束方程为<sup>[2]</sup>:

$$\begin{cases} \min \left\{ \frac{1}{2} \|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2 \right\}, \text{st.} \\ \mathbf{d}_{k, T_r} - \mathbf{X}_{k, T_r} \mathbf{w}_{k+1} = \bar{\mathbf{0}} \end{cases} \quad (4)$$

式中:  $\mathbf{d}_{k, T_r} = [d(k-t_1), d(k-t_2), \dots, d(k-t_r)]^T$ ;  $\mathbf{X}_{k, T_r} = [\mathbf{x}_{k-t_1}, \mathbf{x}_{k-t_2}, \dots, \mathbf{x}_{k-t_r}]^T$ 。同样通过将约束最优化问题转化成无约束最优化问题, 可得此时权值更新表达式为<sup>[2]</sup>:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \mathbf{X}_{k, T_r}^T (\mathbf{X}_{k, T_r} \mathbf{X}_{k, T_r}^T)^{-1} \mathbf{e}_{k, T_r} \quad (5)$$

式中  $\mathbf{e}_{k, T_r} = \mathbf{d}_{k, T_r} - \mathbf{X}_{k, T_r} \mathbf{w}_k$ 。此时目标函数极小表达式为<sup>[2]</sup>:

$$J_{T_r, \min}(\mathbf{w}_{k+1}) = \frac{1}{2} \|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2 + \mathbf{e}_{k, T_r}^T (\mathbf{X}_{k, T_r} \mathbf{X}_{k, T_r}^T)^{-1} \mathbf{e}_{k, T_r} \quad (6)$$

式(2)是输入数据完整时推导出的最优表达式, 而式(5)是在缺少部分输入数据的情况下推导出来的。要使式(5)与式(2)收敛速度最接近, 式(6)与式(3)差别应最小<sup>[2]</sup>。将式(6)中的后半部分写成如下<sup>[2]</sup>:

$$\mathbf{e}_{k, T_r}^T (\mathbf{X}_{k, T_r} \mathbf{X}_{k, T_r}^T)^{-1} \mathbf{e}_{k, T_r} \approx \sum_{i=t_1}^{t_r} \frac{e^2(k-i)}{\|\mathbf{x}_{k-i}\|^2} \quad (7)$$

定义  $e^2(k-i)/\|\mathbf{x}_{k-i}\|^2$  为筛选因子, 由于式(7)的右边项的正定性, 应该选出筛选因子集合中最大的  $R$  项以使式(6)与式(2)差别最小<sup>[2]</sup>。设最大  $R$  项下标为  $\{t_{1opt}, t_{2opt}, \dots, t_{Ropt}\} (t_{1opt} < t_{2opt} < \dots < t_{Ropt})$ , 选出对应的误差项和输入数据有:

$$\mathbf{e}_{k, T_{Ropt}} = [e(k-t_{1opt}), e(k-t_{2opt}), \dots, e(k-t_{Ropt})]^T \quad (8)$$

$$\mathbf{X}_{k, T_{Ropt}} = [\mathbf{x}_{k-t_{1opt}}, \mathbf{x}_{k-t_{2opt}}, \dots, \mathbf{x}_{k-t_{Ropt}}]^T \quad (9)$$

将式(8)和式(9)代入式(5)中即得到了最终的 SR-APA。传统  $L$  阶 APA 每次迭代所需乘法数为  $(L^2+2L)M+L^3+L^2$ , SR-APA 每次迭代所需乘法数为  $((rL)^2+2Lr)M+(rL)^3+(rL)^2$ , 另外多了  $(1-r)LM+L+1$  次乘法和  $L$  次除法用于排序。在  $0.5 \leq r < 1$  时 SR-APA 性能与传统 APA 相比只有微小的改变<sup>[2]</sup>。

## 2 加权部分更新算法

SR-APA 能大量减少 APA 的计算量, 然而该算法在降低 MSE 方面却没有改善, 在计算量相当时比 APA 的 MSE 性能反而要差<sup>[2]</sup>。MSE 理论上为能量均方误差(Energy Mean Squared Error, EMSE)与噪声能量之和, 若要改善算法的 MSE 性能, 即要降低 EMSE。文献[5]给出了计算 APA 额外均方误差 EMSE 的理论式:

$$EMSE = \frac{\mu\sigma_v^2}{(2-\mu)} Tr(R_x) E \left[ \frac{L}{\|\mathbf{x}_k\|^2} \right] \quad (10)$$

从式(10)看出,可以降低数据重用因子  $L$  或者减小步长  $\mu$ 。然而降低  $L$  或者是  $\mu$  的后果是必然会使收敛性能下降<sup>[1,5-6]</sup>。本文提出加权部分更新算法(Weighed SR-APA),可以在降低  $L$  的同时减小 MSE,从而进一步减少 SR-APA 的计算量,提高其性能。

## 2.1 数据加权

首先考虑加权对数据筛选的影响。设  $A^L = \text{diag}(\lambda^0, \lambda^1, \dots, \lambda^{L-1})$ ,  $A^M = \text{diag}(\lambda^0, \lambda^1, \dots, \lambda^{M-1})$ ,  $0 < \lambda \leq 1$ 。对不同时间输入的数据按照指数形式加权得到  $\mathbf{X}_k^A = A^L \mathbf{X}_k A^M$ <sup>[7]</sup>。

设先验误差为  $e_{a,k}(k-i) = \mathbf{x}_{k-i}^T \Delta \mathbf{w}_k = \mathbf{x}_{k-i}^T (\mathbf{w}_0 - \mathbf{w}_k)$ , 则加权后的先验误差可写成:

$$e_{a,k}^A(k-i) = \mathbf{x}_{k-i}^T \text{diag}(\lambda^{i-1}, \lambda^{i-2}, \dots, \lambda^{i-M+1}) \Delta \mathbf{w} \quad (11)$$

在平稳环境中下面的一组式子成立<sup>[5]</sup>:

$$\begin{cases} E\{e_k^{A2}(k-i)\} = E\{e_{a,k}^{A2}(k-i)\} + \sigma_v^2 \\ E\{e_{a,k}^{A2}(k)\} = E\{e_{a,k}^{A2}(k-1)\} = \dots = E\{e_{a,k}^{A2}(k-M+1)\} \end{cases} \quad (12)$$

设  $\sigma_x^2$  为输入数据的功率,很容易得到下面的等式:

$$E\{\|\mathbf{x}_{k-i}^A\|^2\} = E\left\{\sum_{j=i-1}^{i+M-2} \lambda^{k-j} x^2(k-j)\right\} = \sum_{j=i-1}^{i+M-2} \{\lambda^{k-j} E[x^2(k-j)]\} = \frac{\lambda^{i-1}(1-\lambda^{M-1})}{1-\lambda} \sigma_x^2 \quad (13)$$

忽略先验误差与输入数据间的弱相关性<sup>[5]</sup>,对筛选因子取期望并将分式上下展开有:

$$E\left\{\frac{e_k^{A2}(k-i)}{\|\mathbf{x}_{k-i}^A\|^2}\right\} = \frac{E\{e_k^{A2}(k-i)\}}{E\{\|\mathbf{x}_{k-i}^A\|^2\}} = \frac{E\{e_{a,k}^{A2}(k-i)\} + \sigma_v^2}{\frac{\lambda^{i-1}(1-\lambda^{M-1})}{1-\lambda} \sigma_x^2} \quad (14)$$

式(14)中,1)当  $\lambda=1$  时,筛选因子的期望相同,由于输入数据是同分布的,输入数据向量被选出的概率一样,此时 WSR-APA 等效于 SR-APA; 2)若  $\lambda$  非常接近于 0,时间靠前的筛选因子期望要比时间靠后筛选因子的期望大得多,因而往往被选出来的输入数据向量即为最后  $R$  个输入数据向量,此时 WSR-APA 等效于  $R$  阶的传统 APA; 3)若  $\lambda$  取值介于上两者之间,输入时间越靠前,即  $i$  越大,分母越小,筛选因子的期望也就越大,在输入数据同分布情况下被选出的概率也就越大,这样,不加权时输入数据向量被选出来的概率相等,而加权后时间靠后的筛选因子被放大,最优  $R$  集合被约束在时间靠后部分,实际等效于降低了  $L$  的取值。

显然  $\lambda$  越小,最优  $R$  子集选择范围就越小,EMSE 也就越小,极端情况是最优  $R$  子集会只选出最后的  $R$  个输入数据向量。但是往往需要保留对式(6)贡献非常大的输入向量,这样才能充分利用选择性带来的优势,以获得最快的收敛速度,从这个角度出发,  $\lambda$  应该尽量取接近 1 的常数。因而  $\lambda$  应该取接近 1 但小于 1 的一个常数作为折中。

令  $e_{k,T_R}^A = \mathbf{d}_{k,T_R} - \mathbf{X}_{k,T_R}^A \mathbf{w}_k$ , 式(9)变成:

$$e_{k,T_R}^{AT} (\mathbf{X}_{k,T_R}^A \mathbf{X}_{k,T_R}^{AT})^{-1} e_{k,T_R}^A \approx \sum_{i=t_1^A}^{t_R^A} \frac{e_k^{A2}(k-i)}{\|\mathbf{x}_{k-i}^A\|^2} \quad (15)$$

选出集合  $\{e_k^{A2}(k-i) / \|\mathbf{x}_{k-i}^A\|^2\}$  中的最大  $R$  项并保留最大项的下标  $\{t_{1opt}^A, t_{2opt}^A, \dots, t_{Ropt}^A\}$ 。

## 2.2 条件数的考虑

上面的分析是从滤波器本身来考虑,加权改变了数据被选出来的概率。然而 APA 的性能除了与滤波器的结构有关,往往还与输入的数据有关,输入数据相关性越高,输入数据条件数越大,滤波器的收敛性和 MSE 性能都会越差<sup>[8-9]</sup>。实际上可以将式(10)写成式(16)<sup>[8]</sup>,式中条件数定义如式(17)<sup>[10-11]</sup>。

$$EMSE = \frac{\mu\sigma_v^2}{2-\mu} L\chi_E [E(X_i X_i^T)] \tag{16}$$

$$\chi_E(X_i X_i^T) = \frac{1}{L} Tr(X_i X_i^T) Tr[(X_i X_i^T)^{-1}] \tag{17}$$

直接对数据进行加权的问题在于加权后的数据会使条件数大大增加。用式(17)作为条件数的一个量度<sup>[9]</sup>,可以得到图1。图1显示的是白噪声通过一阶AR滤波器后所得数据加权前后的条件数,加权常数为 $\lambda=0.95$ ,1000次平均。

从图1中可以看到原始数据加权后平均条件数会增大不少,这对于算法性能的改进非常不利,条件数的增大不仅会降低收敛速率,从式(16)中可以看出,还会导致EMSE增大。

事实上条件数在对APA的影响主要在于式(3)中求逆过程<sup>[8]</sup>,对原始的APA做一个改进,可以避免加权带来的条件数增大问题,即在求逆时不使用加权后的数据而选用原始数据,即将式(5)改写成:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_k \mathbf{X}_{k,T_R}^T (\mathbf{X}_{k,T_R} \mathbf{X}_{k,T_R}^T)^{-1} \mathbf{e}_{k,T_R}^A \tag{18}$$

式中 $\mathbf{X}_{k,T_R} = [\mathbf{x}_{k-t_1^A}, \mathbf{x}_{k-t_2^A}, \dots, \mathbf{x}_{k-t_R^A}]^T$ 为未加权的数据,但是下标根据加权后的数据选出来,误差表达式如下:

$$\mathbf{e}_{k,T_R}^A = \mathbf{d}_{k,T_R} - \mathbf{X}_{k,T_R}^A \mathbf{w}_k = [e_k(t_{1opt}^A), e_k(t_{2opt}^A), \dots, e_k(t_{Ropt}^A)] \tag{19}$$

### 2.3 可变步长

由于可变步长能够同时取得快的收敛速度和低MSE<sup>[6,12]</sup>,在WSR-APA中引入文献[12]提出的变步长策略:

$$\mathbf{p}(k+1) = \alpha \mathbf{p}(k) + (1-\alpha) \mathbf{X}_{k,T_R}^T (\mathbf{X}_{k,T_R} \mathbf{X}_{k,T_R}^T)^{-1} \mathbf{e}_{k,T_R}^A \tag{20}$$

$$\mu_{k+1} = \frac{\|\mathbf{p}(k)\|^2}{\|\mathbf{p}(k)\|^2 + C} \tag{21}$$

式(20)中 $\alpha$ 为经验值。综上所述得到一个加权部分更新算法(WSR-APA),与SR-APA相比,该算法在加权时多了LM次乘法,而变步长引入的只是一次除法和3次乘法,引入的计算量相对SR-APA计算量而言只算很少一部分。

### 3 算法仿真

本节通过系统识别对文章中所提出的算法进行仿真,并与有关算法进行比较。设未知系统为32阶的FIR系统,由MATLAB随机产生,输入信号为高斯白噪声通过极点为0.9的一阶AR滤波器所得到的信号,SNR为30dB。文中用到的变步长最大步长因子都为1, $\alpha=0.9, C=0.005$ 。WSR-APA的加权常数 $\lambda=0.95$ 。所得到的曲线都经过200次平均。

图2中16阶WSR-APA,SR-APA与8阶的APA计算量相当,但是WSR-APA收敛性能要优于16阶的APA,是4种算法中收敛最快的,MSE也是4种算法中最小的。而SR-APA的收敛性能介于16阶和8阶的APA之间,MSE是4种算法中最差的。在相同阶数下,输入数据选择比例相同时WSR-APA性能要远优于SR-APA。

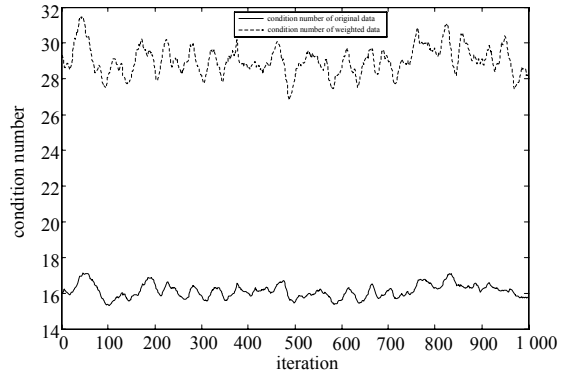


Fig.1 Comparison of condition numbers between original and weighted data  
图1 加权前后条件数的比较

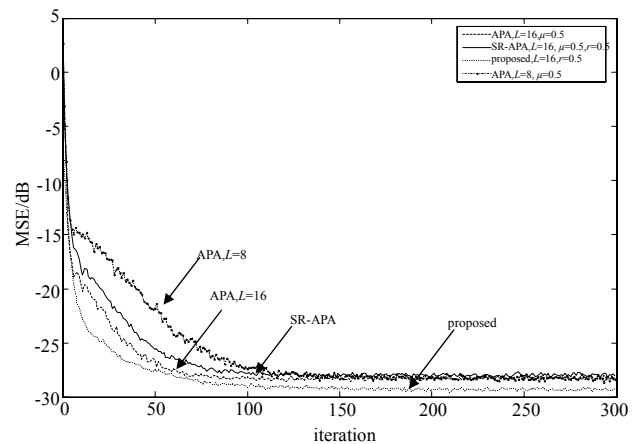


Fig.2 Convergence speed and steady-state MSE performance of APA, SR-APA and proposed algorithm  
图2 APA,SR-APA 和本文算法收敛速度与稳态均方误差性能比较

图2给我们的一个启示是,若WSR-APA与SR-APA收敛速度一样,WSR-APA计算量会远远少于SR-APA。事实上从图3可以看出, $r=0.25$ 时WSR-APA性能要远优于 $r=0.25$ 的SR-APA,其收敛速度与 $r=0.5$ 的SR-APA没有什么区别,MSE还要小。 $L=16$ 时, $r=0.25$ 的WSR-APA的总乘法数约为1760次,而 $r=0.5$ 的SR-APA约为3400,前者只有后者的一半<sup>[2]</sup>。

最后比较变步长和加权带来的影响,如图4。VS-APA为变步长仿射投影算法,VS-SR-APA为只在SR-APA中引入变步长的一种算法。WSR-APA仍是4种算法中性能最好的算法,在收敛性能和稳态误差方面的性能都要比VS-SR-APA好得多,比同阶数VS-APA也要好,可见加权对本文算法起主要作用。另外从图4可得 $r=0.5$ 的WSR-APA的MSE与步长为0.1的APA的一样,收敛速度比此时的APA要快得多。

#### 4 结论

本文在SR-APA的基础上提出了一种新的算法,该算法最大的特点在于对输入数据进行加权,从而改变了SR-APA数据筛选方式,降低了原算法的等效数据重用因子,最终达到以少量的计算成本换取收敛速度和MSE性能的提高。或者说在与SR-APA收敛速度相同时进一步减少SR-APA的计算量,并改善其MSE性能。算法的加权常数 $\lambda$ 在0.9~0.99之间比较合适。另外实验表明,收敛速度相同时,WSR-APA计算量只有SR-APA的50%,而此时在MSE方面,WSR-APA比SR-APA还要好。

#### 参考文献:

- [1] Sayed A H. Fundamentals of Adaptive Filtering[M]. New York:Wiley, 2003.
- [2] Hwang K Y,Song W J. An Affine Projection Adaptive Filtering Algorithm With Selective Regressors[J]. IEEE Tran. Circuit and Systems- II :Express Briefs, 2007,54(1):43-46.
- [3] Dogancay K,Tanrikulu O. Adaptive Filtering Algorithm with Selective Partial updates[J]. IEEE Tran. Circuits Systems-II, Analog Digit, Signal Process, 2001,48(8):762-769.
- [4] Werner S,Apolinario J A,Campos M L R. The Data-selective Constrained Affine Projection Algorithm[C]// ICASSP. Salt Lake City:[s.n.], 2001(6):3745-3748
- [5] Shin H C,Sayed A H. Mean-Square Performance of a Family of Affine Projection Algorithms[J]. IEEE Tran. Signal Processing, 2004,52(1):90-102.
- [6] 倪锦根,李锋. 一种新的变步长最小均方自适应滤波算法[J]. 信息与电子工程, 2008,6(1):10-13. (NI Jinggen,LI Feng. A Novel Variable Step-Size LMS Adaptive Filtering Algorithm[J]. Information and Electronic Engineering, 2008,6(1):10-13.)
- [7] Dai Tao,Shahrrava B,Chen Xiang. A Variable Step-Size Affine Projection Algorithm with a Weighted Projection[C]// IEEE CCECE/C CGEI. Saskatoon:[s.n.], 2005:320-323.
- [8] Lee C W,CHO H,Ban S J,et al. A Derivation of the excess mean square error for affine projection algorithms using the condition number[C]// Control,Automation and Systems, 2007.ICCAS '07. International Conference. Seoul:[s.n.], 2007:1708-1711.
- [9] Rey H,Verge L R,Tressens S,et al. Variable Explicit Regularization in Affine Projection Algorithm: Robustness Issues and Optimal Choice[J]. IEEE Tran. on Signal Processing, 2007,55(5):2096-2109.
- [10] Benesty J,Gansler T. A Recursive Estimation of the Condition Number in the RLS Algorithms[C]// Proc. IEEE International Conference on Acoustics,Speech,and Signal Processing. Montreal:[s.n.], 2005:25-28.

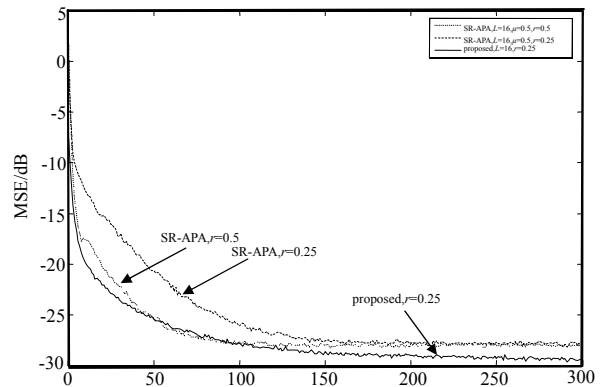


Fig.3 Computation complexity and performance comparison between SR-APA and proposed algorithm

图3 SR-APA与本文算法比较

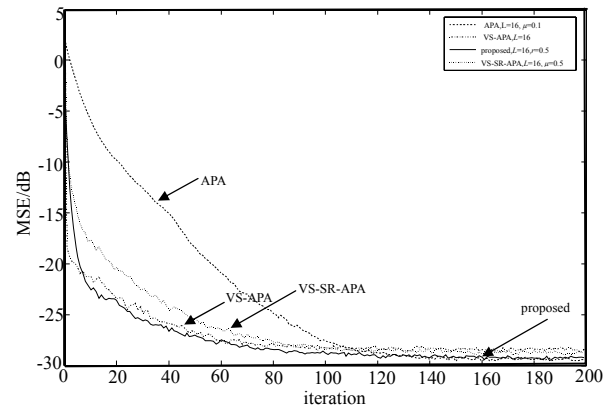


Fig.4 Impact of variable step size and weighting on the proposed algorithm

图4 变步长和加权对本文算法的影响