

A Signal Decorrelation PNLMS Algorithm for Double-Talk Acoustic Echo Cancellation

Kun Pu¹ · Jiashu Zhang¹ · Liangwen Min¹

Received: 13 May 2014 / Revised: 10 April 2015 / Accepted: 11 April 2015 / Published online: 9 May 2015
© Springer Science+Business Media New York 2015

Abstract Acoustic echo cancellation (AEC) in voiced communication systems is used to eliminate the echo which corrupts the speech signal and reduces the efficiency of signal transmission. Usually, the performance of AEC system based on the adaptive filtering degrades seriously in the presence of speech issued from the near-end speaker (double-talk). In typical AEC scenarios, double-talk detector (DTD) must be added to AEC for improving speech quality. One of the main problems in AEC with DTD is that the DTD errors can result in either large residual echo or distorting the near-end input speech. Considering the strong correlation property of speech signals, this paper presents a novel proportionate decorrelation normalized least-mean-square (PDNLMS) adaptive AEC without DTD for echo cancellation as an interesting alternative to the typical AEC with DTDs. Unlike traditional AEC with a DTD, the proposed PDNLMS uses the difference of near-end speech as the residual error to update adaptive echo channel filter during the periods of double-talk, which can efficiently reduce the double-talk influence on the AEC adaptation process. The experimental results show that not only the proposed PDNLMS without DTD illustrate better stability and faster convergence rate, but it is also of a lower steady-state misalignment and better residual signal than current methods with DTDs at a lower computational cost.

✉ Jiashu Zhang
jszhang@home.swjtu.edu.cn

Kun Pu
sabancd@gmail.com

Liangwen Min
lwmin@my.swjtu.edu.cn

¹ The Sichuan Province Key Lab of Signal and Information Processing, Southwest Jiaotong University, Chengdu 610031, People's Republic of China

Keywords Proportionate algorithm · Decorrelation · Acoustic echo cancellation · Double-talk

1 Introduction

Acoustic echo cancellation (AEC) adaptive filters are commonly adopted in voiced communication systems. This kind of filter estimates the impulse response of echo channel between loudspeaker and microphone. A typical double-talk AEC system can be seen in Fig. 1a. Echo signal $y(n)$ is produced by filtering $x(n)$ through an echo channel $w(n)$. The microphone signal consists of echo $y(n)$, near-end speech $s(n)$ and system noise $v(n)$. An adaptive finite impulse response filter $\hat{w}(n)$ is modeled to obtain the replica of echo signal. Echo cancellation is accomplished by subtracting the replica $\hat{y}(n)$ from microphone signal. Even if this question is straightforward, the disturbance of near-end speech brings a great challenge for double-talk echo cancellation. For this reason, a DTD is introduced to sense this condition. Whenever double-talk condition is detected, the filter adaptation is slowed down or completely halted. Nevertheless, the filter may diverge since it works inefficiently during double-talk periods. This problem still remains to be solved.

When no near-end signal is present, several conventional algorithms work well for AEC. The normalized least-mean-square (NLMS) algorithm in [15] achieves both fast convergence rate and low steady-state misadjustment. A number of variable step-size NLMS (VSS–NLMS) algorithms such as [2, 16, 22] are introduced to improve the performance of NLMS. Proportionate NLMS (PNLMS) in [6, 24] improves convergence rate on typical echo paths, and the PNLMS only entails a modest increase in computational complexity compared with NLMS. The affine projection algorithm (APA) and its different versions [10, 19, 21, 25] are attractive choices for acoustic echo cancellation. To improve the performance of APA, Paleologu [19] comes up with a variable step-size APA (VSS–APA). The proposed algorithm aims to recover near-end signal with the error signal and requires no priori information about the acoustic environment. The affine projection sign algorithm (APSA) [21] provides both good robustness and fast convergence. Proportionate APSA [25] designed for achieving a

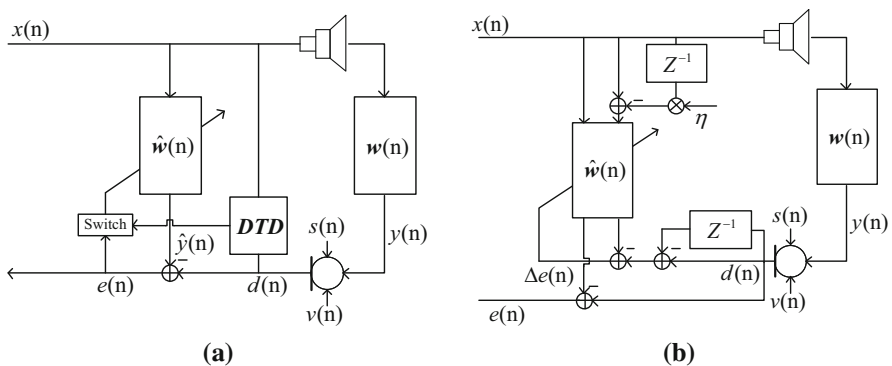


Fig. 1 **a** A typical AEC system. **b** The proposed DTD echo cancellation

better performance under sparse channel is the combination of proportionate approach and APSA. The system [13] dealing with nonlinear acoustic channel is made up of a nonlinear module based on the polynomial Volterra filter and a standard linear module. In [14], a frequency domain postfilter is added to hands-free system. Meanwhile, a psychoacoustically motivated weighting rule is introduced. The new scheme [7] uses the least-mean-square (LMS) algorithm to update the parameters of sigmoid function and the recursive least-square (RLS) algorithm to determine the coefficient vector of the transversal filter. Literature [18] presents an overview about several approaches for the control of step size in adaptive echo cancellation.

However, the presence of near-end speech $s(n)$ disturbs the filter adaptation. As a consequence, the above algorithms fail to work properly. So DTDs in [11, 17, 20, 23] are introduced to stop the filter adaptation when double-talk is detected. Traditional DTD sets the threshold of a certain parameter between the far-end speech and near-end signal. By comparing this parameter with a defined threshold, a decision is made whether the adaptive filter works or should be frozen. Some well-known algorithms deal with the detection of near-end speech. They are mainly based on energy comparison and cross-correlation. The energy-based Geigel algorithm in [5] compares the magnitude of the mixed desire signal $d(n)$ with M most recent samples of $x(n)$. Despite the superiority of being computationally simple, this algorithm does not always perform reliably. An algorithm based on the orthogonality principle is introduced in [3]. According to the idea, a cross-correlation vector between input vector and the scalar microphone output is considered to measure double-talk. In [1, 8, 9], the author proposed a normalized version which is called the normalized cross-correlation (NCR) algorithm. A low complexity version of NCR is presented in [12] for implementation on IP-enabled telephones. However, if the detection threshold is not chosen correctly, the adaptive filter will be influenced and achieves slow convergence rate. So a dynamic one based on signal envelopes is proposed in [23]. The results presented in this literature prove that the accuracy is higher than that in the Geigel algorithm and comparable to the correlation-based methods.

The main objective of this manuscript is to handle double-talk condition without DTDs. To achieve this goal, a signal decorrelation method is introduced to the proposed structure. The decorrelation in this paper is taking the difference of input sequences. In this way, the fluctuation of near-end sequences would not influence the adaptive filter since speech signals are strongly correlated. Adaptive filter keeps refreshing to match the latest echo channel during double-talk periods in the proposed structure. So the proposed algorithm obtains better recovered speech signals compared with its rival DTD algorithms and PNLMS. The paper is organized as follows. In Sect. 2, we present the PDLMS algorithm designed for AEC applications. A brief convergence analysis is given in Sect. 3. Some simulation results are provided in Sect. 4 to support our point of view. Finally, Sect. 5 concludes this work.

2 Proposed Algorithm

In this section, we will state the model in Fig. 1b. Unlike traditional double-talk echo canceller, our proposed structure reduces the correlation of signals and focuses on

minimizing the difference value of $e(n)$ and $e(n - 1)$. This novel model removes the DTD to reduce complexity. The derivation of the algorithm will be presented below. All signals are real-valued.

Let $\mathbf{w}(n)$ denotes the coefficient vector of an unknown echo system with length M , and $\hat{\mathbf{w}}(n)$ represents the replica of $\mathbf{w}(n)$

$$\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{M-1}(n)]^T \quad (1)$$

$$\hat{\mathbf{w}}(n) = [\hat{w}_0(n), \hat{w}_1(n), \dots, \hat{w}_{M-1}(n)]^T \quad (2)$$

Superscript capital letter T denotes transportation. Suppose $y(n)$ is the output signal of an unknown echo path.

$$y(n) = \mathbf{w}^T(n)\mathbf{x}(n) + v(n) \quad (3)$$

$\mathbf{x}(n)$ contains M recent samples of the far-end input signal. $v(n)$ is zero-mean Gaussian noise with variance σ_v^2 . The system noise which corrupts the output of the unknown system is independent of the input sequences.

$$\mathbf{x}(n) = [x(n), x(n - 1), \dots, x(n - M + 1)]^T \quad (4)$$

Let us define posteriori error signals at time n and $n - 1$, respectively

$$e_p(n) = y(n) + s(n) - \hat{\mathbf{w}}^T(n + 1)\mathbf{x}(n) \quad (5)$$

$$e_p(n - 1) = y(n - 1) + s(n - 1) - \hat{\mathbf{w}}^T(n + 1)\mathbf{x}(n - 1) \quad (6)$$

$\hat{\mathbf{w}}(n)$ in (6) is replaced by $\hat{\mathbf{w}}(n + 1)$ approximately. This assumption is reasonable because $\hat{\mathbf{w}}(n)$ changes slowly at every iteration. The desired signal refers to $d(n) = y(n) + s(n)$. Inspired by our previous work [26,27], let us consider the following constrain criteria

$$J(n) = \|\hat{\mathbf{w}}(n + 1) - \hat{\mathbf{w}}(n)\|^2 + \lambda * (e_p(n) - \eta e_p(n - 1)) \quad (7)$$

λ is a Lagrange multiplier. Error difference parameter $\Delta e_p = e_p(n) - \eta e_p(n - 1)$ helps minimize the distinction between $e_p(n)$ and its adjacent sample $e_p(n - 1)$. If we take the partial derivative of (7) with respect to $\hat{\mathbf{w}}(n + 1)$, by setting this derivative to zero the following equations are deduced

$$\frac{\partial J(n)}{\partial \hat{\mathbf{w}}(n + 1)} = 2(\hat{\mathbf{w}}(n + 1) - \hat{\mathbf{w}}(n)) - \lambda(\mathbf{x}(n) - \eta\mathbf{x}(n - 1)) = 0 \quad (8)$$

$$\hat{\mathbf{w}}(n + 1) = \hat{\mathbf{w}}(n) + \frac{\lambda}{2}(\mathbf{x}(n) - \eta\mathbf{x}(n - 1)) \quad (9)$$

Assign $\mathbf{D}(n) = \mathbf{x}(n) - \eta\mathbf{x}(n - 1)$ as the new input signal. $0 \leq \eta < 1$ is called the decorrelation parameter which controls the degree of decorrelation. Obviously, if

$\eta = 0$, our algorithm will turn into NLMS. Combined (9) with $d(n - 1) = \mathbf{x}^T(n - 1)\hat{\mathbf{w}}(n + 1)$ and $d(n) = \mathbf{x}^T(n)\hat{\mathbf{w}}(n + 1)$

$$\begin{cases} d(n) = \mathbf{x}^T(n) (\hat{\mathbf{w}}(n) + \frac{\lambda}{2}\mathbf{D}(n)) \\ d(n - 1) = \mathbf{x}^T(n - 1) (\hat{\mathbf{w}}(n) + \frac{\lambda}{2}\mathbf{D}(n)) \end{cases} \tag{10}$$

$e(n) = d(n) - \mathbf{x}^T(n)\hat{\mathbf{w}}(n)$ and $e(n - 1) = d(n - 1) - \mathbf{x}^T(n - 1)\hat{\mathbf{w}}(n)$ are defined as prior error sequences. From (10), we can obtain the solution of λ as follows

$$\lambda = \frac{2(e(n) - \eta e(n - 1))}{\|\mathbf{D}(n)\|^2} \tag{11}$$

Let $\Delta e(n) = e(n) - \eta e(n - 1)$, so the update equation of filter coefficients can be rewritten as

$$\hat{\mathbf{w}}(n + 1) = \hat{\mathbf{w}}(n) + \frac{\mu \Delta e(n) \mathbf{D}(n)}{\|\mathbf{D}(n)\|^2 + \delta} \tag{12}$$

where δ denotes a variable regularization parameter to avoid the divide-by-zero problem and μ is step-size constant.

The characteristic of echo channel should be taken into consideration before we apply the above algorithm to echo cancellation. Most transmission channels of acoustic echo cancellation are naturally sparse, so the coefficients are zero or close to zero. To accelerate the convergence rate of small coefficients and ensure the quality of recovered near-end speech, we combine the proportionate algorithm in [6] with our method. According to this idea, the update equation of our proportionate decorrelation method can be summarized as

$$\hat{\mathbf{w}}(n + 1) = \hat{\mathbf{w}}(n) + \frac{\mu \Delta e(n) \mathbf{G}(n) \mathbf{D}(n)}{\mathbf{D}^T(n) \mathbf{G}(n) \mathbf{D}(n) + \delta} \tag{13}$$

Compared with (12), $\mathbf{G}(n) = \text{diag}[g_0(n), g_1(n), \dots, g_{M-1}(n)]$ mainly controls the step size of each coefficients to achieve a faster convergence rate. The original definition of $\mathbf{G}(n)$ can be specified in [6]

$$\begin{cases} \gamma_{\min} = \rho \max(\delta_p, |\hat{w}_0(n)|, \dots, |\hat{w}_{M-1}(n)|) \\ \gamma_l(n) = \max(\gamma_{\min}, |\hat{w}_l(n)|) \\ g_l(n) = \gamma_l(n) / \sum_{i=0}^{M-1} \gamma_i(n) \end{cases} \tag{14}$$

δ_p prevents the system from stalling while any filter coefficient equals zero. ρ sets the value of γ_{\min} which controls the minimum adaption rate for each coefficient. For a filter with length $M = 1024$, reasonable δ_p and ρ both may be 0.01.

3 Algorithm Analysis

We employ the tap-weight vector $\hat{\mathbf{w}}(n)$ determined by our proposed adaptive filter to estimate echo system $\mathbf{w}(n)$. Define $\boldsymbol{\varepsilon}(n) = \mathbf{w}(n) - \hat{\mathbf{w}}(n)$ as the mismatch between

$\hat{\mathbf{w}}(n)$ and $\mathbf{w}(n)$. Obviously, this parameter can be adopted to verify the performance and stability of our proposed algorithm. The following equation is obtained combining with (13)

$$\boldsymbol{\varepsilon}(n + 1) = \boldsymbol{\varepsilon}(n) - \frac{\mu \Delta e(n) \mathbf{G}(n) \mathbf{D}(n)}{\mathbf{D}^T(n) \mathbf{G}(n) \mathbf{D}(n) + \delta} \tag{15}$$

Suppose that the input signal $x(n)$ is a zero-mean Gaussian signal with variance σ_x^2 which is independent with $s(n)$ and $v(n)$. This assumption has been widely adopted in NLMS-type analysis [6, 16]. Apparently, $D(n)$ is not an independent Gaussian process for that

$$\begin{aligned} E\{D(n)D(n - 1)\} &= E\{(x(n) - \eta x(n - 1))(x(n - 1) - \eta x(n - 2))\} \\ &= -\eta \sigma_x^2 \neq 0 \end{aligned} \tag{16}$$

Similarly, we can easily obtain $E\{\mathbf{D}^T(n) \mathbf{G}(n) \mathbf{D}(n)\} = (1 + \eta^2) \sigma_x^2$. The denominator of (15) can be regarded as a constant according to [4, 6]. (15) will be rewritten as

$$\boldsymbol{\varepsilon}(n + 1) = \boldsymbol{\varepsilon}(n) - \frac{\mu \mathbf{G}(n) \mathbf{D}(n)}{(1 + \eta^2) \sigma_x^2} \left(\mathbf{D}^T(n) \boldsymbol{\varepsilon}(n) + v_\Delta(n) + s(n) - s(n - 1) \right) \tag{17}$$

$v_\Delta(n)$ applies white Gaussian distribution $(0, 2\sigma_v^2)$ that is independent with $x(n)$. Let $\mathbf{Z}(n) = E\{\boldsymbol{\varepsilon}(n)\}$, if we take the expectation of (17)

$$\mathbf{Z}(n + 1) = \left(\mathbf{I} - \frac{\mu \mathbf{G}(n) \mathbf{R}_{DD}}{(1 + \eta^2) \sigma_x^2} \right) \mathbf{Z}(n) \tag{18}$$

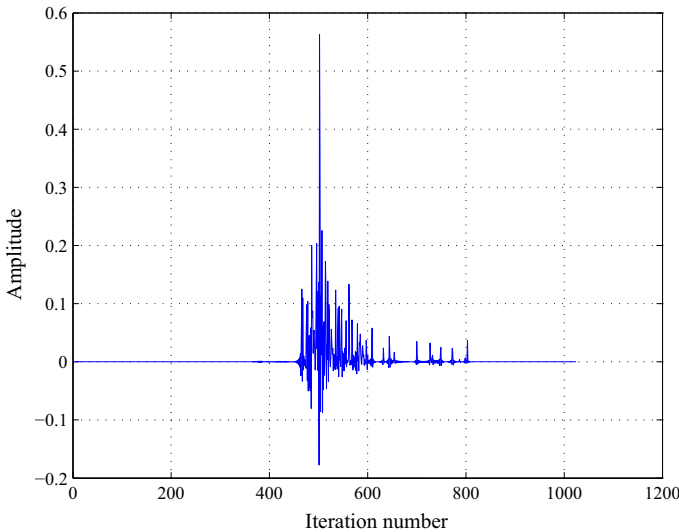


Fig. 2 Acoustic echo channel adopted in the simulation

Table 1 Simulation flow of PDLMS and other comparing algorithms

Method	Algorithm flow
PNLMS [6]	<p>Initialization</p> $\delta_p = \rho = 0.01 \quad \delta = 0.001 \quad \hat{\mathbf{w}}(n) = \mathbf{0} \quad \mathbf{G}(n) = \mathbf{0}$ <p>Update</p> $e(n) = d(n) - \hat{\mathbf{w}}^T(n)\mathbf{x}(n)$ $\gamma_{\min} = \rho \max(\delta_p, \hat{w}_0(n) , \dots, \hat{w}_{M-1}(n))$ $\gamma_l(n) = \max(\gamma_{\min}, \hat{w}_l(n))$ $G_l(n) = \gamma_l(n) / \sum_{i=0}^{M-1} \gamma_i(n), \quad 0 \leq l \leq M - 1$ $\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta}$
Geigel-PNLMS [5]	<p>Initialization</p> $\delta_p = \rho = 0.01 \quad \delta = 0.001 \quad \hat{\mathbf{w}}(n) = \mathbf{0} \quad \mathbf{G}(n) = \mathbf{0}$ <p>Update</p> $\xi(n) = d(n) / \max\{ x(n-1) , \dots, x(n-M) \}$ <p>if $\xi(n) < 0.5$</p> $\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta}$ <p>Otherwise break</p>
EPE-PNLMS [17]	<p>Initialization</p> $\delta_p = \rho = 0.01 \quad \delta = 0.001 \quad \hat{\mathbf{w}}(n) = \mathbf{0} \quad \mathbf{G}(n) = \mathbf{0}$ $\sigma_x^2 = 0 \quad \sigma_d^2 = 0 \quad \lambda = 0.99$ <p>Update</p> $\sigma_x^2 = \lambda\sigma_x^2 + (1-\lambda)x^2(n) \quad \sigma_d^2 = \lambda\sigma_d^2 + (1-\lambda)d^2(n)$ $\xi(n) = \ \hat{\mathbf{w}}(n)\ ^2 \bullet \sigma_x^2 / \sigma_d^2$ <p>if $\xi(n) > 1.1$</p> $\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta}$ <p>Otherwise break</p>
SE-PNLMS [23]	<p>Initialization</p> $\delta_p = \rho = 0.01 \quad \delta = 0.001 \quad \hat{\mathbf{w}}(n) = \mathbf{0} \quad \mathbf{G}(n) = \mathbf{0} \quad \alpha=0.99$ $\gamma=0.05 \quad \beta=0.03 \quad v_x(0) = v_d(0) = v_y(0) = 0 \quad T_{\text{init}}=0.8$ <p>Update</p> $\hat{y}(n) = \hat{\mathbf{w}}^T(n)\mathbf{x}(n)$ $v_x(n) = \alpha v_x(n-1) + (1-\alpha) x(n) $ $v_y(n) = \alpha v_y(n-1) + (1-\alpha) \hat{y}(n) $ $v_d(n) = \alpha v_d(n-1) + (1-\alpha) d(n) $ <p>if $\frac{v_d(n)}{v_x(n)+\gamma} > \frac{v_y(n)}{v_x(n)+\gamma} + \beta$</p> $\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta}$ <p>Otherwise break</p>

Table 1 continued

Method	Algorithm flow
PDDLMS	Initialization $\delta_p = \rho = 0.01 \quad \delta = 0.001 \quad \hat{\mathbf{w}}(n) = \mathbf{0} \quad \mathbf{G}(n) = \mathbf{0} \quad \eta = 0.99$ Update $e(n) = d(n) - \hat{\mathbf{w}}^T(n)\mathbf{x}(n)$ $e(n-1) = d(n-1) - \hat{\mathbf{w}}^T(n)\mathbf{x}(n-1)$ $\mathbf{D}(n) = \mathbf{x}(n) - \eta\mathbf{x}(n-1) \quad \Delta e(n) = e(n) - \eta e(n-1)$ $\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{\Delta e(n)\mathbf{G}(n)\mathbf{D}(n)}{\mathbf{D}^T(n)\mathbf{G}(n)\mathbf{D}(n)+\delta}$

\mathbf{R}_{DD} is the autocorrelation matrix of $\mathbf{D}(n)$, $\mathbf{R}_{DD} = E\{\mathbf{D}(n)\mathbf{D}^T(n)\}$. Suppose that $\mathbf{Y}(n) = (\mathbf{I} - \mu\mathbf{G}(n)\mathbf{R}_{DD}/(\sigma_x^2 + \eta^2\sigma_x^2))$, then

$$\mathbf{Y}(n) = \begin{pmatrix} 1 - \mu g_0(n) & \frac{\mu\eta}{1+\eta^2}g_0(n) & 0 & \cdots & 0 \\ \frac{\mu\eta}{1+\eta^2}g_1(n) & 1 - \mu g_1(n) & \frac{\mu\eta}{1+\eta^2}g_1(n) & \cdots & 0 \\ 0 & \frac{\mu\eta}{1+\eta^2}g_2(n) & 1 - \mu g_2(n) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{\mu\eta}{1+\eta^2}g_{M-1}(n) & 1 - \mu g_{M-1}(n) \end{pmatrix} \tag{19}$$

$\hat{\mathbf{w}}(n) = \mathbf{0}$ and $\mu = 1$ are the initial values, so the equation (18) can be written as

$$\mathbf{Z}(n+1) = \prod_{i=0}^n \mathbf{Y}(i)\mathbf{w}(n) \tag{20}$$

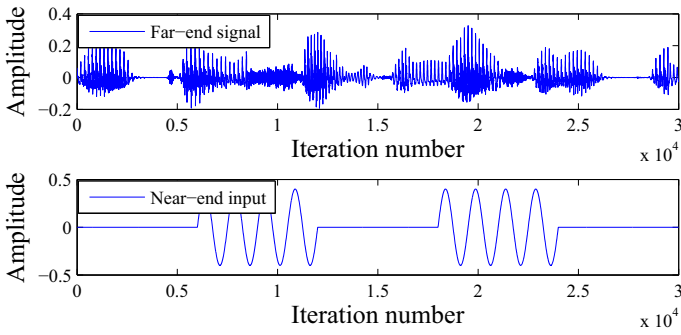
For any $0 \leq \eta < 1$, consider the row norm of matrix $\mathbf{Y}(n)$

$$\|\mathbf{Y}(n)\|_\infty \leq \max\left(1 - \frac{(1-\eta)^2}{1+\eta^2}g_i(n)\right) < 1 \tag{21}$$

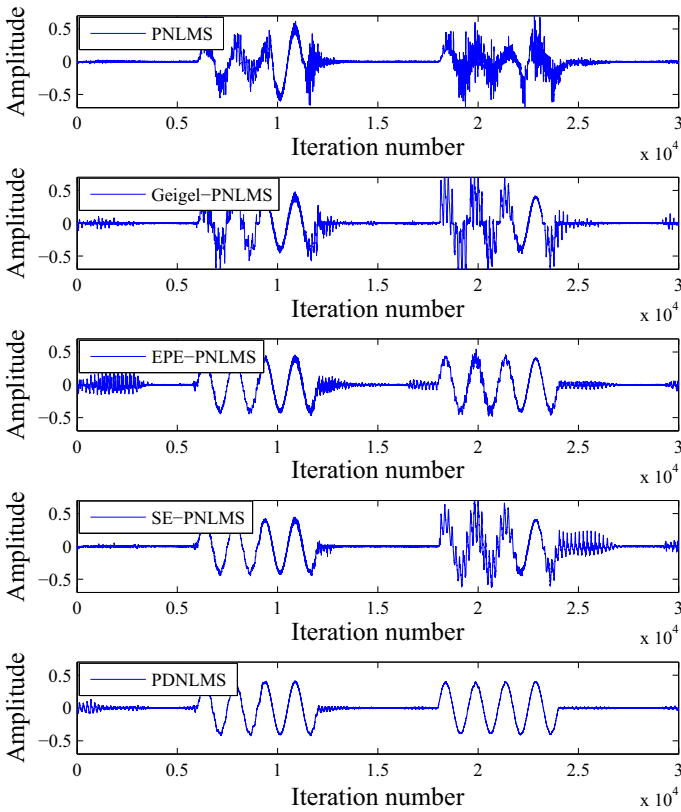
And

$$\|\mathbf{T}(n)\|_\infty = \left\| \prod_{i=0}^n \mathbf{Y}(i) \right\|_\infty \leq \|\mathbf{Y}(0)\|_\infty \|\mathbf{Y}(1)\|_\infty \cdots \|\mathbf{Y}(n)\|_\infty \tag{22}$$

Because $\|\mathbf{Y}(n)\|_\infty$ is always smaller than 1, the limit of $\mathbf{T}(n)$ keeps close to $\mathbf{0}$. Actually, the proposed PDDLMS algorithm converges. According to (21), the parameter η determines the convergence rate and misalignment. When η equals 0, our algorithm turns into PNLMS. This characteristic can be verified in Sect. 4.



(a)



(b)

Fig. 3 a Far-end and near-end input, respectively. b Recovered near-end signals of PNLMS, Geigel-PNLMS, EPE-PNLMS, SE-PNLMS, PDLNLS with sinusoidal near-end input

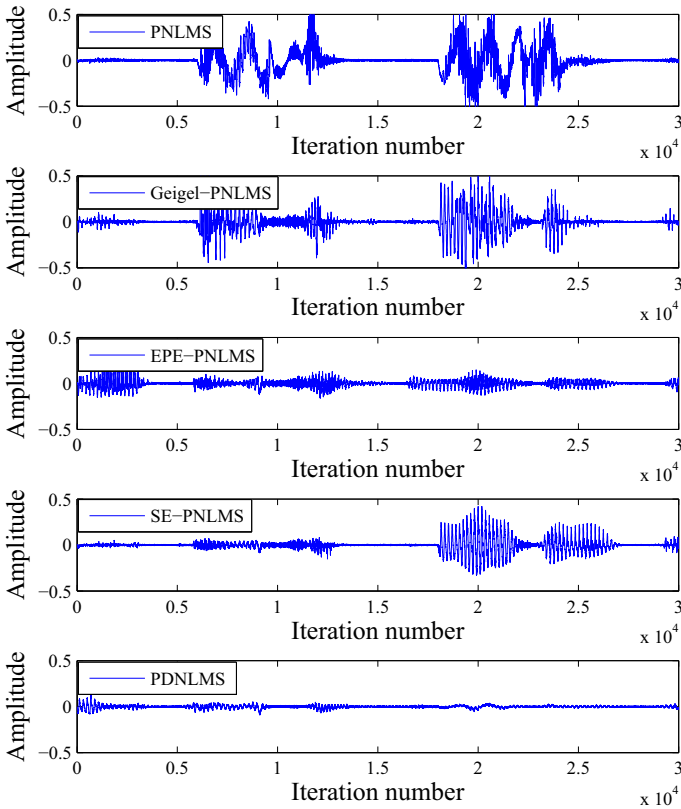


Fig. 4 Residual signals of PNLMS, Geigel-PNLMS, EPE-PNLMS, SE-PNLMS, PDNLMS, respectively, with sinusoidal near-end input

4 Simulations and Results

The simulations of the proposed algorithm are performed in a double-talk scenario. The input signal is either sinusoidal signal generated with a certain frequency or speech sequences. The echo channel with $M = 1024$ can be found in Fig. 2. Two measurements are introduced to evaluate the performance of our PDNLMS algorithm. They are the system misalignment, $20\log(\|w(n) - \hat{w}(n)\| / \|w(n)\|)$ and speech attenuation (SA) during double-talk [20],

$$SA = \frac{1}{K} \sum_{t=1}^K 10 \log \left[\frac{E[s^2(t)]}{E[e^2(t)]} \right] \tag{23}$$

$$\text{Misalignment}(n) = 20 \log \frac{\sum_{k=0}^{N-1} \|w(n) - \hat{w}_i(n - k)\|}{N \|w(n)\|} \tag{24}$$

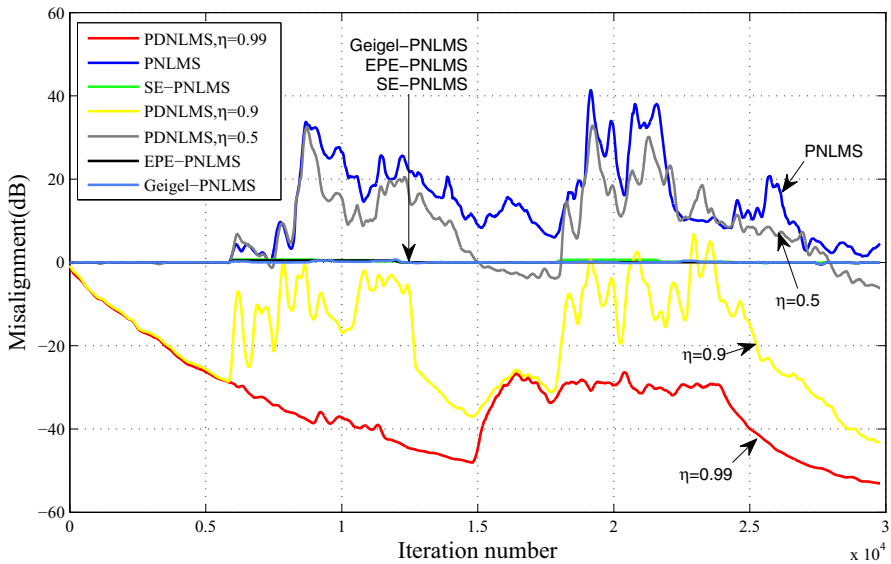


Fig. 5 Misalignment of PNLMS, Geigel-PNLMS, EPE-PNLMS, SE-PNLMS, PDLNLS with sinusoidal near-end input. The far-end input is zero-mean Gaussian AR1 process with variance 0.01 and a pole at 0.99. η equals 0.99, 0.9, 0.5, respectively

where N is the number of samples used in the simulation. $\hat{w}_i(n)$ represents the temporary filter coefficients at i th iteration. K denotes the number of double-talk samples. From (24), the SA of a better recovered signal is always closer to 0.

In this section, we compare PNLMS [6], Geigel-PNLMS [5], EPE-PNLMS [17], the SE-PNLMS [23] with our proposed PDLNLS algorithm. Both echo path and adaptive filters have the length of $M = 1024$. All of our parameter settings can be found in the following Table 1.

4.1 Sinusoidal signal experiments

Firstly, a sinusoidal signal evaluation experiment is considered to test the performance of different systems. $y(n)$ is generated by filtering a speech signal $x(n)$ (sampling rate at 44.1 kHz) through the channel $w(n)$ with length $M = 1024$. The speech comes from a part of an English dialogue. In terms of $s(n)$, sinusoidal signals with digital frequency $f = 1/1500$ and amplitude 0.4 appear every 6000 samples to act as near-end speech. An example of the two signals can be seen in Fig. 3a. An independent Gaussian noise with 25 dB signal-to-noise ratio (SNR) is added to $y(n)$. The step size μ and regularization parameter δ equal to 1 and 0.001, respectively. Recovered near-end signal $e(n)$ is provided with $s(n)$ and $x(n)$ in Fig. 3b.

Figure 3b compares the recovered signals of PNLMS, Geigel-PNLMS, EPE-PNLMS, SE-PNLMS and PDLNLS. It is observed that PNLMS fails to work properly,

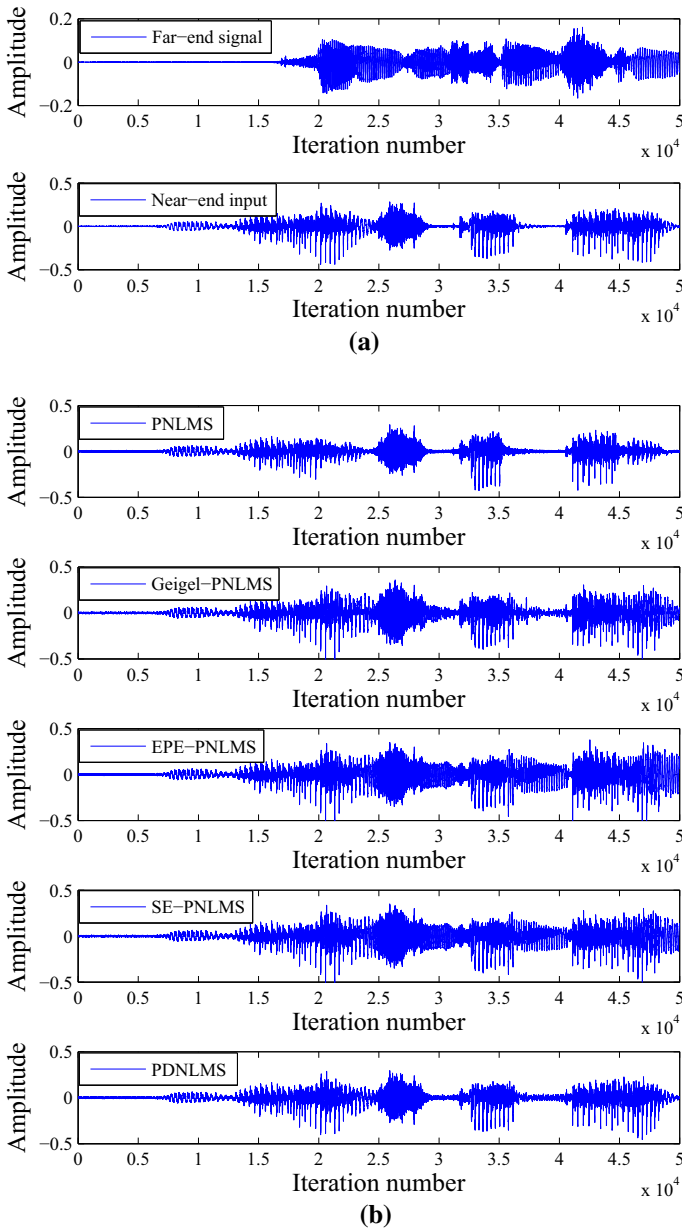


Fig. 6 **a** Far-end and near-end input, respectively. **b** Recovered near-end signals of PNLMS, Geigel-PNLMS, EPE-PNLMS, SE-PNLMS, PDNLMS with near-end speech

while other methods almost draw the outline of $s(n)$. But their results are not stable. PDNLMS produces the best-recovered signal which is the closest to near-end input $s(n)$. The residual signals in Fig. 4 clearly support our conclusion.

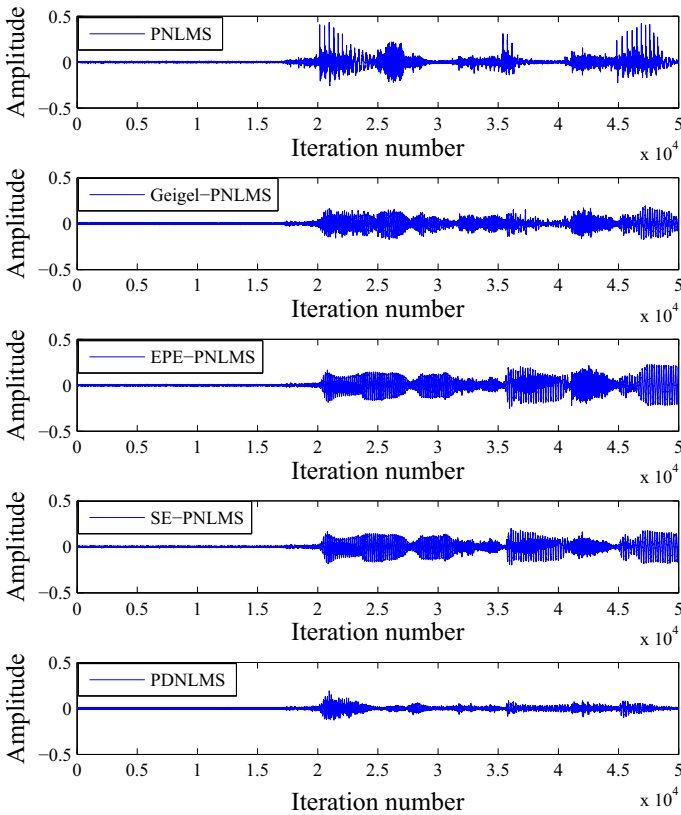


Fig. 7 Residual signals of PNLMs, Geigel-PNLMs, EPE-PNLMs, SE-PNLMs, PDNLMS, respectively, with near-end speech

Secondly, we replace the far-end input by zero-mean Gaussian AR1 process with variance 0.01 and a pole at 0.99. The SNR decreases from 25 to 10 dB between iteration number 15,000 and 24,000, then back to 25 dB after 24,000. All other conditions remain the same with the previous experiment. The misalignment performance of three parameters, where η equals 0.5, 0.9, 0.99, respectively, is provided in Fig. 5. We can see that for different η , the stability and convergence rate get improved while η approaches one. The misalignment curve of PDNLMS with $\eta = 0.99$ outperforms its rival algorithms. In more detail, PNLMs diverges because of the fluctuation of $s(n)$. DTD-based methods achieve similar steady-state misalignment performance because adaptive filter is frozen once double-talk is detected. From Fig. 5, it is observed that our proposed algorithm is immune to the disturbance of near-end signal but a little sensitive to noise. This is because PDNLMS increases the variance of noise. However, the steady-state misalignment of PDNLMS still remains relatively low.

Table 2 Speech attenuation (dB) of PDLMS and other comparing algorithms

Methods	SNR (dB)	Test 1 (20.41 s)	Test 2 (24.94 s)	Test 3 (34.01 s)
PNLMS [6]	15	5.1238	2.5469	2.2840
	20	6.4404	2.9263	2.6682
	25	6.7775	3.0222	2.7500
Geigel-PNLMS [5]	15	2.6516	1.5319	1.1826
	20	4.2795	1.9843	1.6551
	25	4.8193	2.0799	1.7745
EPE-PNLMS [17]	15	-1.4554	0.3177	-0.7634
	20	-0.1217	0.7258	-0.4215
	25	0.4822	0.8393	-0.3519
SE-PNLMS [23]	15	-5.0176	-1.4643	-2.3397
	20	-3.0979	-0.9064	-2.0162
	25	-2.6691	-0.7658	-1.8149
PDLMS	15	-1.0272	-0.0261	-0.2416
	20	0.0958	0.3007	0.0797
	25	0.2985	0.3516	0.1241

4.2 Real Speech Experiments

To investigate the efficiency of our proposed PDLMS in speech applications, the following experiments are employed. Both far-end and near-end sources in our experiment are English recordings sampling at 44.1 KHz. The background is broadcast loudspeaker. The whole system works under the sparse channel in Fig. 2 with the filter length $M = 1024$. White Gaussian noise is added to the echo signal with SNR=25 dB. μ and δ are fixed at 1 and 0.001, respectively, for all algorithms. Signals in Fig. 6a are short versions of the signal sources. Among the outputs in Fig. 6b, our PDLMS fits the near-end input in Fig. 6a best. In Fig. 7, we introduce the residual signal $e(n) - s(n)$ to compare the performance of different algorithms. Obviously, PDLMS still performs much better than the others.

A quantitative assessment of the PDLMS is measured by SA. We test three groups of recordings from two male speakers and a female with the length of 20–35 s. Both near-end and far-end are English recordings. Since the value of K can hardly be verified during a dialogue, we replace it by the length of whole samples. The average SAs of different SNR are recorded in Table 2. We can observe that the SAs of the proposed algorithm are better than those of the rival methods.

5 Conclusion

In this study, the strong correlation of speech signals is taken into consideration to improve the performance of double-talk acoustic echo cancellation. It has been found that adjacent speech samples are quite similar, and the difference value of input

sequence obtains small amplitude. Then, an improved structure based on the signal decorrelation is proposed to handle double-talk AEC without DTD. In this way, the new scheme is able to overcome interferences from near-end. In the proposed algorithm, adaptive filter keeps refreshing during double-talk periods which is different from DTD approaches. Moreover, the computational complexity is reduced by removing the double-talk detector in our algorithm. Simulations show that the architecture is more efficient than those with a DTD.

Nevertheless, our scheme still suffers some problems and need to be further addressed. First, PDLMS increases the variance of noise so the performance may drop when SNR is low. And parameters of the algorithm should be further optimized to work better for low sampling rate signals. Solving these problems should be included in our further work.

Acknowledgments This work was partially supported by National Science Foundation of P.R. China (Grant: 61271341).

References

1. J. Benesty, D.R. Morgan, J.H. Cho, A new class of doubletalk detectors based on cross-correlation. *IEEE Trans. Speech Audio Process.* **8**(2), 168–172 (2000)
2. J. Benesty, H. Rey, L.R. Vega, S. Tressens, A nonparametric vss nlms algorithm. *IEEE Signal Process. Lett.* **13**(10), 581–584 (2006)
3. J.H. Cho, D.R. Morgan, J. Benesty, An objective technique for evaluating doubletalk detectors in acoustic echo cancelers. *IEEE Trans. Speech Audio Process.* **7**(6), 718–724 (1999)
4. H. Deng, M. Doroslovacki, Proportionate adaptive algorithms for network echo cancellation. *IEEE Trans. Signal Process.* **54**(5), 1794–1803 (2006)
5. D.L. Duttweiler, A twelve-channel digital echo canceler. *IEEE Trans. Commun.* **26**(5), 647–653 (1978)
6. D.L. Duttweiler, Proportionate normalized least-mean-squares adaptation in echo cancelers. *IEEE Trans. Speech and Audio Process.* **8**(5), 508–518 (2000)
7. J. Fu, W. Zhu, A nonlinear acoustic echo canceller using sigmoid transform in conjunction with RLS algorithm. *IEEE Trans. Circuits Syst. Express Briefs* **55**(10), 1056–1060 (2008)
8. T. Gansler, J. Benesty, A frequency-domain double-talk detector based on a normalized cross-correlation vector. *Signal Process.* **81**, 1783–1787 (2001)
9. T. Gansler, J. Benesty, The fast normalized cross-correlation double-talk detector. *Signal Process.* **86**, 1124–1139 (2006)
10. S.L. Gay, S. Tavathia, M. Hill, *The Fast Affine Projection Algorithm* (Springer, US, 2000)
11. K. Ghose, V.U. Reddy, A double-talk detector for acoustic echo cancellation applications. *Signal Process.* **80**, 1459–1467 (2000)
12. J.D. Gordy, R.A. Goubran, A low-complexity doubletalk detector for acoustic echo cancellers in packet-based telephony. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 74–77 (2005)
13. A. Guerin, G. Faucon, R. Le Bouquin-jeanns, Nonlinear acoustic echo cancellation based on volterra filters. *IEEE Trans. Speech Audio Process.* **11**(6), 672–683 (2003)
14. S. Gustafsson, R. Martin, P. Jax, P. Vary, A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. *IEEE Trans. Speech Audio Process.* **10**(5), 245–256 (2002)
15. Simon Haykin, *Adaptive Filter Theory* (Prentice-Hall, New Jersey, 2002)
16. H.C. Huang, J. Lee, A new variable step-size NLMS algorithm and its performance analysis. *IEEE Trans. Signal Process.* **60**(4), 2055–2060 (2012)
17. H.K. Jung, N.S. Kim, T. Kim, A new double-talk detector using echo path estimation. *Speech Commun.* **45**, 41–48 (2005)
18. A. Mader, H. Puder, G.U. Schmidt, Step-size control for acoustic echo cancellation-an overview. *Signal Process.* **80**, 1697–1719 (2000)

19. C. Paleologu, J. Benesty, S. Ciochin, A variable step-size affine projection algorithm designed for acoustic echo cancellation. *IEEE Trans. Audio Speech Lang. Process.* **16**(8), 1466–1478 (2008)
20. Y.S. Park, J.H. Chang, Double-talk detection based on soft decision for acoustic echo suppression. *Signal Process.* **90**, 1737–1741 (2010)
21. T. Shao, Y.R. Zheng, J. Benesty, An affine projection sign algorithm robust against impulsive interferences. *IEEE Signal Process. Lett.* **17**(4), 327–330 (2010)
22. H.C. Shin, A.H. Sayed, W.J. Song, Variable step-size NLMS and affine projection algorithms. *IEEE Signal Process. Lett.* **11**(2), 132–135 (2004)
23. G. Szwoch, A. Czyzewski, M. Kulesza, A low complexity double-talk detector based on the signal envelope. *Signal Process.* **88**(11), 2856–2862 (2008)
24. H. Wen, X. Lai, L. Chen, Z. Cai, An improved proportionate normalized least mean square algorithm for sparse impulse response identification. *J. Shanghai Jiaotong Univ.* **18**(6), 742–748 (2013)
25. Z. Yang, Y.R. Zheng, S.L. Grant, Proportionate affine projection sign algorithms for network echo cancellation. *IEEE Trans. Audio Speech Lang. Process.* **19**(8), 2273–2284 (2011)
26. J. Zhang, H.M. Tai, Adaptive noise cancellation algorithm for speech processing, *IECON 2007–33rd Annu. Conf. IEEE Ind. Electron. Soc.* 2489–2492 (2007)
27. J. Zhang, Least mean square error difference minimum criterion for adaptive chaotic noise canceller. *Chin. Phys.* **16**(2), 352–358 (2007)