Combined Regularization Factor for Affine Projection Algorithm Using Variable Mixing Factor

MENGHUA JIANG\textsuperscript{1}, YING GAO\textsuperscript{1}, ZHUORAN CAI\textsuperscript{1}, (Member, IEEE), JINDONG XU\textsuperscript{2}, AND SHIFENG OU\textsuperscript{1}

\textsuperscript{1}School of Physics and Electronic Information, Yantai University, Yantai 264005, China
\textsuperscript{2}School of Computer and Control Engineering, Yantai University, Yantai 264005, China

Corresponding author: Shifeng Ou (ousfeng@ytu.edu.cn)

This work was supported in part by Shandong Provincial Natural Science Foundation under Grant ZR2022MF314, and in part by the National Natural Science Foundation of China under Grant 62072391.

\textbf{ABSTRACT} The affine projection algorithm with a fixed regularization parameter is subject to a compromise concerning the convergence speed and steady-state misalignment. To address this problem, we propose to employ a variable mixing factor to adaptively combine two different regularization factors in an attempt to put together the best properties of them. The selection of the mixing factor is derived by minimizing the energy of the noise-free a posteriori error, and for the sake of suppressing large fluctuations, a moving-average method is designed for updating the mixing factor. Based on a random walk model, we also prove that the proposed mixing factor is as well available for the non-stationary system. The mathematical analysis including the stability performance, steady-state mean square error, and computational complexity are performed. In practice, we compare with the existing related algorithms in system identification and echo cancellation scenarios, the results illustrate that the proposed algorithm outperforms them with notable margins.

\textbf{INDEX TERMS} Adaptive filter, affine projection algorithm, regularization factor, variable mixing factor, system identification.

\section{I. INTRODUCTION}
Adaptive filtering algorithms are attracting considerable interest in numerous signal processing fields \cite{1, 2, 3, 4, 5}. And due to the simple structure and implementation, the least mean square (LMS) and normalized least mean square (NLMS) algorithms have already gained broad application \cite{6, 7}. Since these two LMS-type algorithms will deteriorate in convergence speed when the input signals are correlated, the affine projection algorithm (APA) was proposed by Ozeki K \cite{8} to ensure a superior convergence rate. However, matrix inversion is inevitable in the classical APA, which may give rise to a singularity. Considering practical reasons, a diagonal matrix obtained through multiplying an identity matrix by a positive constant $\delta$, which is called the regularization factor, is generally added to the matrix requiring inverse to prevent numerical problems \cite{6, 7}. With emphasizing the presence of the regularization parameter in mind, such an algorithm is named the regularized affine projection algorithm (R-APA). Despite this fact, there is still a conflicting objective between convergence speed and steady-state misalignment for the R-APA with a fixed regularization factor.

To tackle the aforementioned issue, extensive research in recent years has concentrated on the variable regularization factor (VRF) strategy. Early on, a pseudo-optimal regularization factor for the APA was carried out in \cite{9}, in which the input signals tend to be pre-whitened additionally. An alternative adaptation of the regularization matrix was developed by the normalized stochastic gradient in \cite{10}. Furthermore, by setting the a posteriori error equal to the noise variance, \cite{11} put forward a variable regularization factor...
APA (VR-APA). Based on the premise that the statistical characteristic of the noise is fully considered, [12] proposed a novel VRF formulation to reduce the steady-state misalignment, but the exponential scaling factor estimation method introduced several parameters and excessively relied on some a priori knowledge. Later on, an improved version of the pseudo APA (PAPA) [13] whose regularization factor varies was derived to overcome the problems of the PAPA. And by minimizing the $l_2$-norm of the a priori error vector, [14] proposed a variable regularization affine projection sign algorithm (VR-APSA) to improve the convergence performance. Traditionally, the already proposed VRF methods can alleviate the tradeoff between convergence speed and steady-state error to some extent, but the challenge of high steady-state errors is still present in some environments.

Not only the VRF schemes but also some adaptive combination approaches have generated much research attention to improve the performance of the filters, which can perform at least as well as the best of the components. Earlier attempts pertaining to the combination of APA were reported in [15] and [16]. The construct of the convex combination was first articulated in [15], which can effectively solve the contradiction problem. Nevertheless, there is still a distinctive two-stage convergence phenomenon in practical simulations. [16] achieved the convex combination between two different R-AP filters and briefly introduced two improvement methods. Indeed, the improved algorithms shown in [16] intend to relieve the two-stage convergence, there is much room to accelerate the convergence speed. Such a combination scheme has also been extensively applied to some specific adaptive filters [17], [18], [19], [20], [21]. In these works, it was well demonstrated that the convex combination of adaptive filters can provide a promising way to improve performance.

In this paper, to take full advantage of the fast convergence of the small regularization factor and low steady-state misalignment of the large regularization factor, we propose a combined regularization factor for affine projection algorithm (CRF-APA), which adaptively associates two different regularization factors with a variable mixing factor that is obtained by minimizing the energy of the noise-free a posteriori error. Subsequently, to refrain from large fluctuation, the mixing factor is streamlined in a moving-average method. If the variance of the noise is unknown, an efficient learning strategy is applied to ascertain its estimate. For the integration of the algorithm, we exemplify that the mixing factor is similarly available for a non-stationary environment, which is concluded from a random walk model. To illustrate the above, we have done theoretical analyses containing the stability performance, steady-state mean square error (MSE), and computational complexity. Finally, experiments in system identification and echo cancellation scenarios serve as further validation of the proposed algorithm.

The rest of the paper is organized as follows. In section 2, the R-AP recursion is concisely described. And section 3 illustrates the proposed algorithm with aspects of the stationary system and non-stationary system. Some performance analyses were carried out in Section 4. Section 5 exhibits numerical simulation results aimed at assessing the performance of the proposed algorithm. Ultimately, section 6 summarizes the conclusion.

II. BRIEF VIEW OF THE R-APA

We consider that the input signal matrix $\mathbf{U}(n) = [\mathbf{u}(n), \mathbf{u}(n-1), \ldots \mathbf{u}(n - K + 1)]$ is filtered through the unknown system $\omega_o$ to get the desired output vector $\mathbf{d}(n)$.

$$\mathbf{d}(n) = \mathbf{U}^T(n) \omega_o + \mathbf{v}(n)$$

where $n$ is the time index, and $\mathbf{v}(n)$ denotes the noise vector. Besides, in this paper, $L$ is the filter length, and $K$ is the projection order.

And a priori error vector can be written as

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{U}^T(n) \omega(n)$$

where $\omega(n)$ is the estimate of $\omega_o$ at iteration $n$.

From [6] and [7], through the Lagrange multiplier and gradient descent method, the update equation of the conventional R-APA is given by

$$\omega(n + 1) = \omega(n) + \mu \mathbf{U}(n) \left( \mathbf{U}^T(n) \mathbf{U}(n) + \delta \mathbf{I} \right)^{-1} \mathbf{e}(n)$$

where $\mu$ accounts for the step size, and $\delta$ denotes the regularization factor which is generally used to provide numerical stability. Furthermore, as far as the R-APA is concerned, a small regularization factor equates to a large step size, hence the algorithm can converge fast, while also generating a large steady-state misalignment. Instead, a large regularization factor is comparable to a small step size, which will result in a relatively small misalignment only but poor convergence. For that reason, we are keen to explore a novel scheme that can combine the respective merits of different regularization factors to improve the behavior of the algorithm.

III. PROPOSED CRF-APA

A. STATIONARY SYSTEM

It’s worth noting that the value of $\mu$ should meet $0 < \mu < 1$ or $1 < \mu < 2$ for ensuring convergence stability [22]. And the former has a lower steady-state error and the filter converges fastest when the $\mu$ equals 1. Therefore, as is done in [23] and [24], we assume the step size $\mu$ to be one.

Substituting the fixed regularization factor $\delta$ with the combination $\delta(n) = t(n)\delta_1 + (1-t(n))\delta_2$ in (3) yields

$$\omega(n + 1) = \omega(n) + \mu \mathbf{U}(n) \left( \mathbf{U}^T(n) \mathbf{U}(n) + \delta(n) \mathbf{I} \right)^{-1} \mathbf{e}(n)$$

where $\delta_1$ and $\delta_2$ are constants ($\delta_1 < \delta_2$), and $t(n)$ represents the variable mixing factor to dynamically adjust the combined regularization factor $\delta(n)$. 

Utilizing the weight-error vector $\tilde{w}(n) = \omega_0 - \omega(n)$ for (4) gives rise to

$$\tilde{w}(n + 1) = \tilde{w}(n) + U(n) \left(U^T(n) U(n) + \delta(n) I \right)^{-1} e(n)$$  \hspace{1cm} (5)

Pre-multiplying (5) by $U^T(n)$ and using $e(n) = e_a(n) + v(n)$, we obtain

$$e_p(n) = e_a(n) - U^T(n) U(n) \left(U^T(n) U(n) + \delta(n) I \right)^{-1} e(n)$$

$$= \left(I - U^T(n) U(n) \left(U^T(n) U(n) + \delta(n) I \right)^{-1}\right) e_a(n) - U^T(n) U(n) \left(U^T(n) U(n) + \delta(n) I \right)^{-1} v(n)$$  \hspace{1cm} (6)

where $e_a(n) = U^T(n) \tilde{w}(n)$ denotes noise-free a priori error vector, and $e_p(n) = U^T(n) \tilde{w}(n + 1)$ denotes noise-free a posteriori error vector.

Before going a step further, we need to make some assumptions:

**Assumption 1:** The noise is independent and identically distributed and statistically independent of the regression matrix $U(n)$ and $U(n)$ is statistically independent of $e(n)$ at steady-state [24]

**Assumption 2:** The expectation of the ratio between two random variables can be similar to the ratio of their expectations. [23], [25]

Taking the square and expectation of (6), since some factors disappear under Assumption 1, we arrive at

$$E \left[ e_p^2(n) \right] = E \left[ e_a^T(n) \left(I - U^T(n) U(n) \left(U^T(n) U(n) + \delta(n) I \right)^{-1}\right) e_a(n) \right]$$

$$+ E \left[ v^T(n) \left( U^T(n) U(n) \left(U^T(n) U(n) + \delta(n) I \right)^{-1} \right) v(n) \right]$$  \hspace{1cm} (7)

To simplify the expression (7), we perform the eigenvalue decomposition of the matrix $U^T(n) U(n)$:

$$U^T(n) U(n) = Q(n) \Lambda(n) Q^T(n)$$  \hspace{1cm} (8)

where $\Lambda(n)$ denotes a diagonal matrix formed by the eigenvalues of $U^T(n) U(n)$, and $\Lambda(n) = \text{diag}(\lambda_1(n), \lambda_2(n), \ldots, \lambda_K(n))$. In addition, $Q(n)$ represents a matrix whose columns are the corresponding eigenvectors of $U^T(n) U(n)$. Based on (8), we get

$$\left(U^T(n) U(n) + \delta(n) I \right)^{-1} = Q(n) \left( \Lambda(n) + \delta(n) I \right)^{-1} Q^T(n)$$  \hspace{1cm} (9)

Substituting (9) into (7) yields

$$E \left[ e_p^2(n) \right] = E \left[ e_a^T(n) Q(n) \left( \frac{\delta(n)}{\Lambda(n) + \delta(n) I} \right)^2 Q^T(n) e_a(n) \right]$$

$$+ E \left[ v^T(n) Q(n) \left( \frac{\Lambda(n)}{\Lambda(n) + \delta(n) I} \right)^2 Q^T(n) v(n) \right]$$  \hspace{1cm} (10)

Taking the first-order derivative of the right-hand side of (10) with respect to $t(n)$ in $\delta(n)$ and setting the result to zero leads to

$$E \left[ e_p^T(n) Q(n) \text{diag} \left\{ \frac{\delta(n)}{\Lambda(n) + \delta(n) I} \right\} Q^T(n) e_a(n) \right]$$

$$= E \left[ v^T(n) Q(n) \text{diag} \left\{ \frac{\Lambda(n)}{\Lambda(n) + \delta(n) I} \right\} Q^T(n) v(n) \right]$$  \hspace{1cm} (11)

where $1 \leq k \leq K$. And just like [12] and [24], the eigenvalues of $U^T(n) U(n)$ can be approximated as

$$\lambda_k(n) \approx L \sigma_k^2(n)$$  \hspace{1cm} (12)

where $L \sigma_k^2(n)$ can be regarded as the average of all eigenvalues and $\sigma_k^2(n)$ is the estimate of the variance of the input signal $u(n)$, which can be calculated by

$$\sigma_k^2(n) \approx \frac{\text{Tr} \left[U^T(n) U(n)\right]}{LK}$$  \hspace{1cm} (13)

where $\text{Tr}\{\cdot\}$ denotes the trace of the matrix. Then substituting (12) and (13) into (11) and invoking Assumption 2, we can get

$$E \left[ \|e_a(n)\|^2 \right] t(n) (\delta_1 - \delta_2 + \delta_2) = E \left[ \|v(n)\|^2 \right] L \sigma_k^2(n)$$  \hspace{1cm} (14)

Therefore, the updating formula of the mixing parameter $t(n)$ is

$$t(n) = \frac{E \left[ \|v(n)\|^2 \right] L \sigma_k^2(n) - \delta_2 E \left[ \|e_a(n)\|^2 \right]}{E \left[ \|e_a(n)\|^2 \right] (\delta_1 - \delta_2)}$$

$$= \frac{K \sigma_v^2(n) L \sigma_k^2(n) - \delta_2 \sigma_{e_a}^2(n)}{\sigma_{e_a}^2(n) (\delta_1 - \delta_2)}$$  \hspace{1cm} (15)

where $E[\|v(n)\|^2] = K \sigma_v^2(n)$ and $E[\|e_a(n)\|^2] = \sigma_{e_a}^2(n)$.

Next, we approximate the power of noise-free a priori error $\sigma_{e_a}^2(n)$ in (15) by the time average of the squares of $e_a(n)$.

$$\sigma_{e_a}^2(n) = \sigma_1 \sigma_{e_a}^2(n - 1) + (1 - \sigma_1) \|e_a(n)\|^2$$  \hspace{1cm} (16)

where $\sigma_1$ is a forgetting factor, which satisfies $0.9 < \sigma_1 < 1$.

The vector $e_a(n)$ can be recovered from the noisy a priori error vector $e(n)$ by using the shrinkage denoising method described in [26] and we can obtain the solution by the following steps (17-20).

$$q = \sqrt{K \sigma_v^2}$$  \hspace{1cm} (17)
\[
a_{o}(n) = D^{T}e(n)
\]
\[
a(n) = \text{sign}(a_{o}(n)) \odot g(n)
\]
\[
e_{a}(n) = Da(n)
\]

where \(D\) is an orthogonal matrix, \(\odot\) denotes the element-wise product, \(g(n) = [g_{1}(n), g_{2}(n), \ldots, g_{K}(n)]\) whose element satisfies

\[
g_{r}(n) = \max \{|a_{o,r}(n) - q_{0}|, 1 \leq r \leq K\}
\]

where \(a_{o,r}(n)\) is the \(r\)-th element of \(a_{o}(n)\).

In order to more smoothly update the mixing factor \(t(n)\), we apply the time-averaging method to (15).

\[
t(n) = a_{2}t(n - 1) + (1 - a_{2}) \min \left\{ \frac{K\sigma_{r}^{2}L\sigma_{e}^{2}(n) - \delta_{2}\sigma_{e}^{2}(n)}{\sigma_{e}^{2}(n)(\delta_{1} - \delta_{2})}, t(n - 1) \right\}
\]

The time-average method in (22) can also guarantee the monotonic decrease of \(t(n)\), nevertheless, it may yield negative values when the algorithm arrives at the steady-state. To overcome this shortcoming, we provide an adjustment as

\[
t(n) = \begin{cases} 
a_{2}t(n - 1) + (1 - a_{2}) & \text{if } t(n) > 0 \\
\min \left\{ \frac{K\sigma_{r}^{2}L\sigma_{e}^{2}(n) - \delta_{2}\sigma_{e}^{2}(n)}{\sigma_{e}^{2}(n)(\delta_{1} - \delta_{2})}, t(n - 1) \right\} & \text{otherwise} 
\end{cases}
\]

where \(t_{\text{min}}\) is a positive constant about zero [27]. Notably, if \(t(n) < 0\), \(t(n)\) will equal \(t_{\text{min}}\), so as to avoid any appearance of negative values.

### B. NONSTATIONARY SYSTEM

It is supposed that the unknown weight vector \(\omega_{o}(n)\) at iteration \(n\) has the following random walk model [28]

\[
\omega_{o}(n + 1) = \omega_{o}(n) + \varphi(n)
\]

where \(\varphi(n)\) is a random accumulation, which is independent and identically distributed and independent of \(\{\omega_{o}(0), \omega(0)\}, \{x_{j}| j \geq 0\}, \{t_{v}||j \geq 0\} \). Each element of \(\varphi(n)\) is of zero mean with the variance \(\sigma_{\varphi}^{2}\). And then (5) can be reformulated as

\[
\bar{\omega}(n + 1) = \bar{\omega}(n) - U(n)\left(U^{T}(n)U(n) + \delta(n)I\right)^{-1}\times e(n) + \varphi(n)
\]

Similarly, pre-multiplying (25) by \(U^{T}(n)\) yields

\[
e_{p}(n) = e_{a}(n) - U^{T}(n)U(n)\left(U^{T}(n)U(n) + \delta(n)I\right)^{-1}e(n) + \varphi(n)
\]

\[
= \left(\mathbf{I} - U^{T}(n)U(n)\left(U^{T}(n)U(n) + \delta(n)I\right)^{-1}\right)e_{a}(n) - U^{T}(n)u(n)\left(U^{T}(n)U(n) + \delta(n)I\right)^{-1}v(n) + U^{T}(n)\varphi(n)
\]

Since the additional term does not have any dependency on the other variables in (25), squaring both sides of (26) and taking the expectation, we have

\[
E\left[e_{p}^{2}(n)\right] = E\left[e_{a}^{2}(n)\right] - \left(\mathbf{I} - U^{T}(n)U(n)\left(U^{T}(n)U(n) + \delta(n)I\right)^{-1}\right)e_{a}(n)
\]

\[
+ \left(U^{T}(n)U(n)U^{T}(n)U(n) + \delta(n)I\right)^{-1}e_{a}(n)
\]

The term \(E\left[\varphi^{T}(n)U^{T}(n)U(n)\varphi(n)\right]\) is involved with the mixing factor \(t(n)\), and then it will vanish after differentiating with regard to \(t(n)\). Consequently, solving (27) gives the same result as further processing (7), thus we conclude that the proposed algorithm is also applicable to the nonstationary system.

### C. PRACTICAL CONSIDERATIONS

1) ESTIMATION OF THE NOISE VARIANCE

The estimation of noise variance is of great importance to the proposed CRF-APA. And if the variance of the background noise is unknown in advance, we can find its estimate using the following strategy [29].

\[
\sigma_{e}^{2} = \sigma_{e}^{2}(n) - E\left[\tilde{\omega}^{T}(n)u(n)u^{T}(n)\tilde{\omega}(n)\right] 
\]

\[
\approx \sigma_{e}^{2}(n) - \frac{r_{ue}(n)r_{ue}(n)}{\sigma_{e}^{2}(n)}
\]

\[
\sigma_{e}^{2}(n) = \alpha_{3}\sigma_{e}^{2}(n - 1) + (1 - \alpha_{3})e^{2}(n)
\]

\[
r_{ue}(n) = \alpha_{3}r_{ue}(n - 1) + (1 - \alpha_{3})u(n)e(n)
\]

with \(\alpha_{3}\) meaning the forgetting factor close to one.

Based on (13), (28)-(30), the variance \(\sigma_{e}^{2}(n)\) can be obtained.

2) RESET METHOD FOR A SYSTEM SUDDEN CHANGE

Due to the monotonically decreasing behavior of the mixing factor \(t(n)\), it must be re-initialized each time the unknown system is changed, otherwise the CRF-APA would lose its tracking capability. For addressing this issue, the reset method illustrated in [23] is used as a reference and modified, as summarized in Table 1.

### IV. PERFORMANCE ANALYSIS

#### A. STABILITY PERFORMANCE

To guarantee the stability of the proposed algorithm and avoid any appearance of negative values, we need to set \(t_{\text{min}}\). In this section, we need to utilize a posteriori error vector \(e(n)\), which
TABLE 1. The reset method.

Initialization $e_{avg}(0) = e^2(I)$, $\rho$ (user defined)
for each iteration $n$
$e_{add}(n)=e(n)$
if $\log_{10}(e_{add}(n)/e_{avg}(n)) > \rho$
$t(n)=1, \sigma_{e}^2(n)=0, \sigma_{e_{add}}^2(n)=0$
end
end
$e_{avg}(n) = \beta e_{avg}(n-1) + (1-\beta) e^2(n)$

is defined as

$$e(n) = d(n) - U^T(n)\omega(n+1)$$  \hspace{1cm} (31)

Substituting (4) into (31) yields

$$e(n) = \left(I - U^T(n)U(n) \left(U^T(n)U(n) + \delta(n)I\right)^{-1}\right)e(n)$$  \hspace{1cm} (32)

and

$$\|e(n)\|^2 - \|e(n)\|^2$$

$$= e^T(n) \left(I - \left(I - U^T(n)U(n) \left(U^T(n)U(n) + \delta(n)I\right)^{-1}\right)^T \right) \times \left(I - \left(I - U^T(n)U(n) \left(U^T(n)U(n) + \delta(n)I\right)^{-1}\right)^T \right)e(n)$$  \hspace{1cm} (33)

From (33), the result is selected to satisfy $E[||e(n)||^2] \leq E[||e(n)||^2]$ (with equality only when $e(n) = 0$). Using the result in (8), the following holds:

$$\left(I - U^T(n)U(n) \left(U^T(n)U(n) + \delta(n)I\right)^{-1}\right)^T$$

$$\times \left(I - U^T(n)U(n) \left(U^T(n)U(n) + \delta(n)I\right)^{-1}\right)e(n)$$

$$= I - 2Q(n)\Lambda(n)(\Lambda(n) + \delta(n)I)^{-1}Q^T(n)$$

$$+ Q(n)(\Lambda(n)\Lambda(n) + \delta(n)I)^{-1}\right)^2Q^T(n)$$  \hspace{1cm} (34)

Furthermore, we deduce that

$$\left(I - \left(I - U^T(n)U(n) \left(U^T(n)U(n) + \delta(n)I\right)^{-1}\right)^T \times \left(I - \left(I - U^T(n)U(n) \left(U^T(n)U(n) + \delta(n)I\right)^{-1}\right)^T \right)e(n)$$

$$= Q(n)\Lambda(n)(\Lambda(n) + \delta(n)I)^{-1}$$

$$\times \left(2I - \Lambda(n)(\Lambda(n) + \delta(n)I)^{-1}\right)Q^T(n)$$  \hspace{1cm} (35)

In order to satisfy that (35) is positive-definite, the term $2I - \Lambda(n)(\Lambda(n) + \delta(n)I)^{-1}$ should be positive. Taking an expectation, we obtain

$$E \left[ 2I - \Lambda(n)(\Lambda(n) + \delta(n)I)^{-1} \right]$$

$$= 2I - E \left[ \Lambda(n)(\Lambda(n) + \delta(n)I)^{-1} \right]$$

$$= 2I - \left[ \text{diag} \left\{ \frac{\lambda_k(n)}{\lambda_k(n) + t(n)(\delta_1 - \delta_2 + \delta_2)} \right\} \right]$$

$$> 0$$  \hspace{1cm} (36)

As a result, the lower bound of the mixing factor for the proposed algorithm can be expressed as

$$t(n)_{\text{min}} > \max_{1 \leq k \leq K} \left( E \left[ \frac{\lambda_k(n)}{1} + 2\delta_2 \right] \right)$$  \hspace{1cm} (37)

B. STEADY-STATE MSE

Based on Assumption 1, Shin et al. [30] have indicated that the formula concerning the steady-state MSE of the APA in a stationary system is

$$MSE = \sigma_v^2 + \frac{\mu\sigma_u^2}{2\eta_u - \mu\alpha_u}$$  \hspace{1cm} (38)

where $\alpha_u$, $\eta_u$, and $A(n)$ are, respectively, calculated by

$$\alpha_u \triangleq \text{Tr}\left( S \cdot E \left[ A(n) \right] \right)$$  \hspace{1cm} (39)

$$\eta_u \triangleq \text{Tr}\left( S \cdot E \left[ \left( U^T(n)U(n) + \delta I \right)^{-1} \right] \right)$$  \hspace{1cm} (40)

$$A(n) = \left( U^T(n)U(n) + \delta I \right)^{-1}U^T(n)U(n)$$

$$\times \left( U^T(n)U(n) + \delta I \right)^{-1}$$  \hspace{1cm} (41)

with $S = \text{diag}\{1, 0, \ldots, 0\}$. Particularly, in [31], the steady-state MSE of the APA with a small regularization factor $\delta$ and large filter length $L$ has been simplified as

$$MSE = \sigma_v^2 + \frac{\mu\sigma_u^2K}{2 - \mu}$$  \hspace{1cm} (42)

However, the key to the proposed CRF-APA is to work on the effect of $\delta$ on the APA regardless of the size of the value. Therefore, we need to allow for a large $\delta$ and the evaluation of the mixing factor $t(n)$ [32].

Based on (8), performing the eigenvalue decomposition of (41) yields

$$A(n) = \frac{Q(n)(\Lambda(n) + \delta(n)I)^{-1}}{\Lambda(n)} \frac{Q(n)}{\Lambda(n)}$$

$$= \frac{Q(n)diag}{Q(n)} \left\{ \frac{\lambda_k(n)}{(\lambda_k(n) + t(n)(\delta_1 - \delta_2 + \delta_2)^2)} \right\} Q^T(n)$$  \hspace{1cm} (43)

Exploiting an evolutionary method, the exact mathematical expectation in (38) can be substituted with an instantaneous value. And the instantaneous values of $\alpha_u$ and $\eta_u$ are described as

$$\alpha_u = \text{Tr}\left( S \cdot [A(n)] \right) = \frac{L\sigma_u^2}{\left( L\sigma_u^2 + t(n)(\delta_1 - \delta_2) + \delta_2 \right)^2}$$  \hspace{1cm} (44)
The algorithm performance, which is defined as
\[ \text{NMSD} (n) = 10 \log_{10} \left( \frac{\| \omega (n) - \omega_v \|^2}{\| \omega_v \|^2} \right) \]

Besides, we assume that the noise variance is known for [8], [22], and [33]. In the simulations, the signal-to-noise ratio (SNR) is stated as
\[ \text{SNR} = 10 \log_{10} \frac{E \left[ (u^T(n) \omega)^2 \right]}{E \left[ v^2 (n) \right]} \]

### A. SYSTEM IDENTIFICATION APPLICATION

Here, simulations are designed for system identification. We randomly generate the unknown system \( \omega_v \) by \( \| \omega_v \|^2 = 1 \). The length of the unknown filter is set to 32, and we assume the adaptive filter has the same number of taps. The colored input signals are generated by filtering a white Gaussian random sequence through a first-order system \( G(z) = 1 / (1 - 0.9z^{-1}) \). What’s more, the projection order is set to \( K = 2 \) and all the learning curves are generated over 100 independent trials to avoid losing the generality.

First of all, we evaluate the performance of the proposed CRF-APA compared with that of the conventional R-APAs, CR-APA, ICR-APA, VR-APAs, and En-APA in a stationary environment with SNR = 30dB. Fig. 1 plots the NMSD learning curves of the above algorithms and Fig. 2 displays the curve of the variable mixing factor of CRF-APA. All the parameters are tuned so that the referred algorithms have both fast convergence and small steady-state error, which are stated in Table 3. As can be seen in Fig. 1, the proposed algorithm achieves both a fast convergence rate of the small regularization factor and high steady-state misalignment of the large regularization factor, while VR-APAs have poor performance in terms of steady-state misalignment and the ICR-APA based on CR-APA converges more slowly than the other algorithms. Notably, the En-APA has a lower steady-state error than VR-APAs but converges slower sometimes. And from Fig. 2, we can visualize that the variation of the mixing factor of CRF-APA decreases from 1 to the minimum value as expected, which can ensure that CRF-APA converges quickly and also achieves a low steady-state misalignment.

Additionally, in adaptive algorithms, tracking capability is also a crucial issue. Figs. 3 and 4 plot the NMSD learning

### C. COMPUTATIONAL COMPLEXITY

Table 2 exhibits the iteration-wise computational complexity values of the traditional R-APA, CR-APA [16], ICR-APA [16], VR-APAs [11, 12], En-APA [33], and CRF-APA in terms of the total number of multiplications, additions, and comparisons. It can be seen from the data that the CR-APA and ICR-APA have high computational costs because they are bound to update two filters with different regularization factors in the meantime. Due to only one filter involved, the En-APA is relatively easy to implement. Because of the presence of the mixing factor, the CRF-APA just mentions only one filter actually. And the shrinkage denoising method just introduces a trivial computational burden. Nevertheless, since the matrix trace is imperative in dealing with the input signals in (13), which causes a greater computational cost, the CRF-APA doesn’t have a significant advantage in the calculation. The trace of the matrix is also adopted by the VR-APAs. More detailed studies on computational complexity are illustrated in [7] and [32].

### V. SIMULATION RESULTS

In this section, we present the simulation comparison of the proposed algorithm and other corresponding state-of-the-art competing methods in two applications, namely, system identification and echo cancellation scenarios. And we use the normalized mean-square deviation (NMSD) to measure the algorithm performance, which is defined as
\[ \text{NMSD} (n) = 10 \log_{10} \left( \frac{\| \omega (n) - \omega_v \|^2}{\| \omega_v \|^2} \right) \]

Notably, compared with (42), the result of (47) is not only related to the variance \( \sigma_v^2(n) \), but also associated with the input signal and the larger regularization factor.
TABLE 3. Tuning parameters of the algorithms.

| Algorithm       | Parameters                        |
|-----------------|-----------------------------------|
| R-APAs          | $\delta_1=0.001$, $\delta_2=3000$, $\mu=1$ |
| CR-APA          | $\mu=10$                          |
| ICR-APA         | $\mu=10$                          |
| VR-APA in [11]  | $\alpha=0.984(\alpha=1-1/\eta L)$ |
| VR-APA in [12]  | $\alpha=0.984$, $\gamma=3$       |
| En-APA          | $K_{max}=8$, $\rho=0.1$, $\delta=0.001$, $\gamma=0.97$ |
| CRF-APA         | $\delta_1=0.001$, $\delta_2=3000$, $\alpha_1=0.998$, $\alpha_2=0.9999$, $\alpha_3=0.99995$, $\lambda_{min}=10^{-4}$ |

Fig. 1. NMSD learning curves of the R-APAs, CR-APA, ICR-APA, VR-APAs, En-APA, and CRF-APA in a stationary environment, with SNR = 30dB.

Fig. 2. The variation curve of the mixing factor of CRF-APA in a stationary environment, with SNR = 30dB.

curves in a non-stationary environment simulation and the variable mixing factor curve of CRF-APA in this case, respectively. We assume that the unknown system impulse responses are regenerated abruptly at iteration $3.5 \times 10^4$. And the reset parameters are set as $\rho=1$, $\beta=0.95$. As shown in Fig.3, the CRF-APA is able to track the changed system quickly, neither losing the convergence speed nor increasing the misalignment. Fig. 4 also exhibits the desired behaviors. It can be noticed that the value of the mixing factor is 1 in the initial period of convergence or when the system changes and finally decreases to the minimum value at the steady-state. These simulations amply indicate that the CRF-APA is highly robust against environmental changes and does not involve a time-consuming parameter adjustment process.

Lastly, aimed at the impact of different noise environments on the algorithms, we simulate similar experiments again for a non-stationary system under different noise settings. Fig. 5 depicts the NMSD learning curves of the aforesaid algorithms in the simulations carried out in the same environment as that shown in Fig. 3, except for SNR = 20dB. Fig. 6 examines the effect of the algorithms under the SNR is 10dB, and for better visualization, we change the value of the step size to $\mu = 0.2$ (for R-APAs, VR-APAs and CRF-APA) and $\mu = 0.06$ (for En-APA). Besides, we turn the settings of the regularization factors into 0.1 and 1100. It is clearly observed that the CRF-APA still provides better performance in the case of SNR = 20dB and SNR = 10dB. These simulations validate the conjectures made by us regarding the behavior of the proposed algorithm irrespective of the noise level.

B. ECHO CANCELLATION APPLICATION

In this subsection, the performance of the proposed CRF-APA is tested for echo cancellation and compared with the R-APAs, CR-APA, ICR-APA, VR-APAs, and En-APA again. As shown in Fig. 7, we set the length of the echo path to 128,
and the same length is used for the adaptive filters. The input speech signal is sampled at 8kHz as illustrated in Fig. 8. And an independent white Gaussian noise signal is added to the output of the echo path with SNR = 30dB. All the simulation results are obtained after 30 independent trials. What’s more, just like [19] and [34], we add another performance evaluation metric called echo return loss enhancement (ERLE), which is described in (50). The higher the value, the better the algorithm’s ability to eliminate echoes.

\[
ERLE = 10 \log_{10} \frac{E[|d|^2(n)]}{E[e^2(n)]} \tag{50}
\]

Fig. 9 reveals convergence curves of the CRF-APA together with the other algorithms with the speech sequence as input. In this case, \(\alpha_1 = 0.998\), \(\alpha_2 = 0.9999\), \(\alpha_3 = 0.999995\), \(\delta_1 = 0.01\), \(\delta_2 = 50\), \(K = 2\), and other parameters are selected to obtain the optimal performance for all algorithms. As can be seen, compared with other algorithms, the proposed CRF-APA has the fastest convergence and lowest steady-state error, which confirms the effectiveness of the mixing factor with aspects of astringency and error performance under the echo cancellation scenario. We can also conclude similar conclusions in Fig.10, which depicts the ERLE comparison of the algorithms. And Table 4 exhibits the corresponding mean ERLE values. An improvement in ERLE...
Mean ERLE values of the R-APAs, CR-APA, ICR-APA, VR-APAs, En-APA and CRF-APA with the input signal is speech.

TABLE 4. Mean ERLE values of the R-APAs, CR-APA, ICR-APA, VR-APAs, En-APA and CRF-APA with the input signal is speech.

| Algorithms | ERLE(dB) | Algorithms | ERLE(dB) |
|------------|----------|------------|----------|
| δ=0.01     | 19.99    | δ=50       | 17.83    |
| CR-APA     | 20.07    | ICR-APA    | 20.53    |
| VR-APA in [11] | 19.02 | VR-APA in [12] | 19.08 |
| En-APA     | 19.08    | CRF-APA    | 21.03    |

is observed in the CRF-APA compared to the existing algorithms, and reflects the superiority of the proposed algorithm.

VI. CONCLUSION

In this paper, we presented a novel method for deriving the variable mixing factor that allows incorporating the strengths of two separate regularization factors, which intends to improve the performance in the convergence rate and steady-state error. Considering the impact of environmental noise, the mixing factor was deduced by minimizing the energy of noise-free a posteriori error. And using the instantaneous estimation and shrinkage denoising method, a practical approximation of the optimal mixing factor has been shown as well. When the environment changes suddenly, the algorithm demonstrated strong tracking behavior and ongoing effectiveness. Through extensive analyses and simulations, our proposed algorithm proved to have competitive convergence speed and produce small steady-state misalignment for both system identification and echo cancellation scenarios, as compared with other available algorithms.

REFERENCES

[1] F. Huang, J. Zhang, and S. Zhang, “Combined-step-size affine projection sign algorithm for robust adaptive filtering in impulsive interference environments,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 63, no. 5, pp. 493–497, May 2016, doi: 10.1109/TCSII.2015.2505067.

[2] X. Wang and J. Han, “Affine projection algorithm based on least mean fourth algorithm for system identification,” IEEE Trans. Signal Process., vol. 58, pp. 11930–11938, 2020, doi: 10.1109/TSP.2020.2966038.

[3] K. Kumar, M. L. N. S. Karthik, S. S. Bhattacharjee, and N. V. George, “Affine projection champernowne algorithm for robust adaptive filtering,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 69, no. 3, pp. 1947–1951, Mar. 2022, doi: 10.1109/TCSII.2021.3124563.

[4] J. Shin, B. Y. Park, W. I. Lee, J. Yoo, and J. Cho, “A novel normalized subband adaptive filter algorithm based on the joint-optimization scheme,” IEEE Access, vol. 10, pp. 9868–9876, 2022, doi: 10.1109/ACCESS.2022.3143136.

[5] D. G. Tiglea, R. Candido, and M. T. M. Silva, “A variable step size adaptive algorithm with simple parameter selection,” IEEE Signal Process. Lett., vol. 29, pp. 1774–1778, 2022, doi: 10.1109/LSP.2022.3195662.

[6] S. Haykin, Adaptive Filter Theory. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[7] A. H. Sayed, Fundamentals of Adaptive Filtering. Hoboken, NY, USA: Wiley, 2003.

[8] K. Ozeki and T. Umeda, “An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties,” Electron. Commun. Jpn., vol. 67, no. 5, pp. 19–27, 1984, doi: 10.1002/ecja.4400670503.

[9] V. Myllyla and G. Schmidt, “Pseudo-optimal regularization for affine projection algorithms,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Budapest, Hungary, May 2002, pp. 427–430, doi: 10.1109/ICASSP.2002.5745003.

[10] Y. S. Choi, H. C. Shin, and W. J. Song, “Adaptive regularization matrix for affine projection algorithm,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 54, no. 12, pp. 1087–1091, Dec. 2007, doi: 10.1109/TCSII.2007.906931.

[11] W. Yin and A. S. Mehr, “A variable regularization method for affine projection algorithm,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 57, no. 6, pp. 476–480, Jun. 2010, doi: 10.1109/TCSII.2010.2048355.

[12] L. Shi and L. Yun, “Variable regularization parameter for affine projection algorithm based on energy of the noise-free a posterior error vector signal,” Acta Electron. Sinica, vol. 43, no. 1, pp. 7–12, 2015, doi: 10.3969/j.issn.0372-2112.2015.01.002.

[13] S. J. Ban, C. W. Lee, and S. W. Kim, “Adaptive regularization parameter for pseudo affine projection algorithm,” IEEE Signal Process. Lett., vol. 16, no. 5, pp. 382–385, May 2009, doi: 10.1109/LSP.2009.2016479.

[14] Y.-R. Chien, “Variable regularization affine projection sign algorithm in impulsive noisy environment,” IEICE Trans. Fundam. Electron. Commun. Comput. Sci., vol. 102, no. 5, pp. 725–728, May 2019, doi: 10.1587/transfun.E102.A.725.

[15] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, “Mean-square performance of a convex combination of two adaptive filters,” IEEE Trans. Signal Process., vol. 54, no. 3, pp. 1078–1090, Mar. 2006, doi: 10.1109/TSP.2005.863126.

[16] K.-H. Kim, Y.-S. Choi, S.-E. Kim, and W.-J. Song, “Convex combination of affine projection filters with individual regularization,” in Proc. 23rd Int. Tech. Conf. Circuits Syst., Comput. Communications (ITC-CSCC), Jul. 2008, pp. 901–904.
[17] L. Shi, Y. Lin, and X. Xie, “Combination of affine projection sign algorithms for robust adaptive filtering in non-Gaussian impulse interference,” *Electron. Lett.*, vol. 50, no. 6, pp. 466–467, Mar. 2014, doi: 10.1049/el.2013.3997.

[18] S. H. Kim, J. J. Jeong, G. Koo, and S. W. Kim, “Robust convex combination of affine projection-type algorithms using an impulsive noise indicator,” *Signal Process.*, vol. 129, pp. 33–37, Dec. 2016, doi: 10.1016/j.sigpro.2016.05.034.

[19] Y.-R. Chien and J. Li-You, “Convex combined adaptive filtering algorithm for acoustic echo cancellation in hostile environments,” *IEEE Access*, vol. 6, pp. 16138–16148, 2018, doi: 10.1109/ACCESS.2018.2804298.

[20] M. Scarpiniti, D. Comminiello, and A. Uncini, “Convex combination of spline adaptive filters,” in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, A Coruña, Spain, Sep. 2019, pp. 1–5, doi: 10.23919/EUSIPCO.2019.8903134.

[21] Y.-R. Chien, C.-H. Yu, and H.-W. Tsao, “Affine-projection-like maximum correntropy criteria algorithm for robust active noise control,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 2253–2266, 2022, doi: 10.1109/TASLP.2022.3190720.

[22] S. G. Sankaran and A. A. L. Beex, “Convergence behavior of affine projection algorithms,” *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1086–1096, Apr. 2000, doi: 10.1109/78.827542.

[23] L. Shi, H. Zhao, W. Wang, and L. Lu, “Combined regularization parameter for normalized LMS algorithm and its performance analysis,” *Signal Process.*, vol. 162, pp. 75–82, Sep. 2019, doi: 10.1016/j.sigpro.2019.04.014.

[24] H. Rey, L. R. Vega, S. Tressens, and J. Benesty, “Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice,” *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096–2109, May 2007, doi: 10.1109/TSP.2007.893197.

[25] H.-S. Lee, S.-H. Yim, and W.-J. Song, “$Z$-proportionate diffusion LMS algorithm with mean square performance analysis,” *Signal Process.*, vol. 131, pp. 154–160, Feb. 2017, doi: 10.1016/j.sigpro.2016.06.011.

[26] Z. A. Bhotto and A. Antoniou, “A family of shrinkage adaptive-filtering algorithms,” *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1689–1697, Apr. 2013, doi: 10.1109/TSP.2012.2236831.

[27] H.-C. Huang and J. Lee, “A new variable step-size NLMS algorithm and its performance analysis,” *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2055–2060, Apr. 2012, doi: 10.1109/TSP.2011.2181505.

[28] C. H. Lee and P. Park, “Scheduled-step-size affine projection algorithm,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 9, pp. 2034–2043, Sep. 2012, doi: 10.1109/TCSI.2012.2185298.

[29] M. A. Iqbal and S. L. Grant, “Novel variable step size NLMS algorithms for echo cancellation,” in *Proc. IEEE Int. Conf. Acoust.*, *Speech Signal Process.*, Las Vegas, NV, USA, Mar. 2008, pp. 241–244, doi: 10.1109/ICASSP.2008.4517591.

[30] H.-C. Shin and A. H. Sayed, “Mean-square performance of a family of affine projection algorithms,” *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 90–102, Jan. 2004, doi: 10.1109/TSP.2003.820077.

[31] S.-E. Kim, S.-J. Kong, and W.-J. Song, “An affine projection algorithm with evolving order,” *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 937–940, Nov. 2009, doi: 10.1109/LSP.2009.2027638.

[32] T. Zhang, H.-Q. Jiao, and Z.-C. Lei, “Individual-activation-factor memory proportionate affine projection algorithm with evolving regularization,” *IEEE Access*, vol. 5, pp. 4939–4946, 2017, doi: 10.1109/ACCESS.2017.2682918.

[33] J. Shin, J. Kim, T.-K. Kim, and J. Yoo, “An enhanced affine projection algorithm based on the adjustment of input-vector number,” *Entropy*, vol. 24, no. 3, p. 431, Mar. 2022, doi: 10.3390/e24030431.

[34] S. Burra and A. Kar, “Performance analysis of an improved split functional link adaptive filtering algorithm for nonlinear AEC,” *Appl. Acoust.*, vol. 176, May 2021, Art. no. 107863, doi: 10.1016/j.apacoust.2020.107863.

**MENGHUA JIANG** received the bachelor’s degree from Shandong Normal University, Jinan, China, in 2020. She is currently pursuing the master’s degree with Yantai University, Yantai, China. Her research interests include adaptive filtering algorithm, sparse signal processing, and machine learning.

**YING GAO** received the Ph.D. degree from Jilin University, Jilin, China, in 2008. She is currently an Associate Professor and a Master Tutor with the School of Physics and Electronic Information, Yantai University, Yantai, China. Her research interests include signal and information processing.

**ZHUORAN CAI** (Member, IEEE) received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, in 2013. He is currently an Associate Professor with the School of Physics and Electronic Information, Yantai University, Yantai, China. His research interests include deep learning methods, cognitive radio, and compressed sensing.

**JINDONG XU** received the Ph.D. degree from Beijing Normal University. He is currently an Associate Professor and a Master Tutor with the School of Computer and Control Engineering, Yantai University, Yantai, China. His research interest includes image processing.

**SHIFENG OU** received the Ph.D. degree from Jilin University, Jilin, China, in 2008. He is currently a Professor and a Master Tutor with the School of Physics and Electronic Information, Yantai University, Yantai, China. His research interests include speech signal processing, adaptive filtering algorithm, and blind signal processing.