Affine Projection Algorithm with Variable Projection Order

Reza Arablouei
Institute for Telecommunications Research
University of South Australia
Mawson Lakes, Australia

Kutluyl Doğançay
School of Electrical and Information Engineering
University of South Australia
Mawson Lakes, Australia

Abstract—Increasing the projection order in the affine projection adaptive filtering algorithm speeds up the convergence but also increases the steady-state misalignment. To address this unfavorable compromise, we propose a new affine projection algorithm with a variable projection order. This algorithm adaptively changes the projection order according to the estimated variance of the filter output error. The error variance is estimated using the exponential window and moving averaging techniques and employing a variable forgetting factor. Simulations demonstrate that the new algorithm provides fast initial convergence and low steady-state misalignment without necessarily trading off one for the other in addition to a significant reduction in average computational complexity.

Keywords—adaptive filtering; affine projection algorithm; variable projection order; error variance estimation

I. INTRODUCTION

The affine projection algorithm (APA) is a stochastic-gradient-based adaptive filtering algorithm that possesses improved convergence rate in comparison with the least mean squares (LMS) algorithm, in particular when the input signal is highly correlated [1]. However, the APA incurs a higher steady state error than the LMS algorithm. Considerable research has been devoted to cope with this intrinsic drawback of the APA, which imposes a trade-off between high initial convergence rate and low steady-state mean square error (MSE). Some approaches focus on controlling the step-size [2]–[4] or regularization parameter [5], [6] of the algorithm, while others aim to control its projection order [7]–[9] or to select optimal input regressors [9]–[12]. In [7], authors propose a set-membership APA with variable data-reuse factor (SM-AP vdr algorithm) that changes the projection order during adaptation utilizing the information provided by the data-dependent step-size of the set-membership filter. The APA with data selective method proposed in [10] decides whether to use new input vectors or not based on their impact on estimated magnitude of the condition number of the input autocorrelation matrix. A similar technique is used in [11] utilizing E-norm instead of L₂-norm to estimate the condition number. Both algorithms impose excessive computational overhead for a moderate performance improvement. The APA with selective regressors (SR-APA) is proposed in [12], where a criterion is developed to decide the rank of each input regressor vector in contributing to the convergence and a fixed number of the highest rank input regressors are selected at each iteration to update filter coefficients. This algorithm is effective when the input signal is extremely correlated and a large number of input regressors are used in the selection process. The APA with dynamic selection of input vectors (DS-APA) proposes to select a dynamic number of input regressors based on a criterion devised to maximize gradient of the mean square deviation [9]. The APA with evolving order (E-APA) [8] tunes the projection order using approximation of the theoretical estimate given in [13] for the steady-state MSE of the APA.

In this paper, we propose a new affine projection algorithm that changes the projection order at each iteration in accordance with the adaptation state. The new algorithm provides an appreciably improved convergence performance. We also introduce a new technique for estimating the variance of the filter output error that has an enhanced tracking capability.

II. AFFINE PROJECTION ALGORITHM

Let us consider data \( \{d(n)\} \) that arise from the model

\[
d(n) = w^*X(n) + v(n)
\]

where \( w \) is an unknown \( L \times 1 \) column vector, \( v(n) \) is the measurement noise, \( X(n) \) is the \( L \times 1 \) input regressor vector

\[
x(n) = [x(n), x(n - 1), \ldots, x(n - L + 1)]^T
\]

and the superscripts \( T \) and * denote matrix transpose and conjugate transpose, respectively. Taking \( w(n) \) as an estimate for \( w \) at iteration \( n \), the affine projection algorithm with a projection order of \( k \) calculates \( w(n) \) via

\[
w(n + 1) = w(n) + \mu X(n)(X^*(n)X(n) + \delta I_k)^{-1} e^*(n)
\]

where \( \mu \) is the step-size, \( \delta \) is the regularization parameter, \( I_k \) is a \( k \times k \) identity matrix,

\[
X(n) = [x(n), x(n - 1), \ldots, x(n - k + 1)],
\]

\[
d(n) = [d(n), d(n - 1), \ldots, d(n - k + 1)],
\]

and

\[
e(n) = d(n) - w^*(n)X(n).
\]

Defining weight error vector as \( \tilde{w}(n) \equiv w_n - w(n) \), we have

\[
\tilde{w}(n + 1) = \tilde{w}(n) - \mu X(n)(X^*(n)X(n) + \delta I_k)^{-1} e^*(n).
\]

Neglecting \( \delta \), gradient of the mean square deviation (MSD) is defined as

\[
\Delta(n) \equiv E[\|\tilde{w}(n)\|^2] - E[\|\tilde{w}(n + 1)\|^2]
\]

\[
= 2\mu \text{Re} \left( E\left[ e(n)(X^*(n)X(n))^{-1}X^*(n)\tilde{w}(n)\right]\right)
\]

\[
- \mu^2 E\left[ e(n)(X^*(n)X(n))^{-1} e^*(n)\right]
\]

This work was supported in part by a Commonwealth Scientific and Industrial Research Organisation (CSIRO) scholarship.
where \( E[\cdot] \) is the expectation operator, \( \| \cdot \| \) stands for the Euclidean norm of a vector and \( \text{Re}(\cdot) \) for the real part of a complex quantity.

### III. Problem Statement

It is known that for a given APA adaptive filtering scenario, the larger the projection order is, the faster the APA converges but to a larger steady-state MSE (or coefficients misalignment) \([13]–[18]\). The improvement in convergence rate can be attributed to the decorrelating property of the APA \([14]\). Alternatively, it can be ascribed to the APA’s feature of producing direction vectors that are strongly correlated with the weight error vector \([15]\). However, these enhancements come at the expense of amplifying energy of the background noise and consequently increasing the steady-state MSE. This is supposedly caused by the APA’s linear but time-varying pre-whitening filter and is aggravated by larger projection orders \([14], [16]\).

Fig. 1 plots MSD gradient of the APA, \( \Delta(n) \), with different projection orders \((k = 1 \text{ to } 4)\) in a typical system identification set-up \((L = 16, \mu = 0.3, \text{SNR} = 10 \text{ dB}, \text{zero-mean unit-variability Gaussian input, and ensemble-averaging over } 10^4 \text{ independent trials})\). It is seen that at the early stages of adaptation, larger projection orders result in larger \( \Delta(n) \) but by approaching the steady state, larger projection orders produce smaller \( \Delta(n) \). Accordingly, the circles in Fig. 1 indicate the time instances at which the use of a smaller projection order leads to a larger \( \Delta(n) \) and consequently faster convergence. These observations corroborate that in order to achieve the fastest convergence, \( \Delta(n) \) should be maximized during the entire filtering operation. This leads to MSD undergoing the largest descent at each iteration and also guarantees achievement of the lowest possible steady-state MSE \([4]\). To maximize \( \Delta(n) \), we may define the following optimization problem:

\[
\kappa_p(n) = \underset{1 \leq k \leq k_{\text{max}}}{\text{arg max}} \; \Delta(n) \tag{1}
\]

where \( \kappa_p(n) \) is the optimum time-varying projection order and is bounded between 1 and a predefined maximum value \( k_{\text{max}} \).

It is worthwhile to note that in \([2]\) the projection order is considered fixed and a time-varying step-size rule is derived based on maximization of \( \Delta(n) \). Here, we consider a fixed step-size and let the projection order change.

### IV. APA with Variable Projection Order

Any exact analytical solution for the optimization problem of \((1)\) will be clearly either impracticable or too costly to implement in real-world applications. In addition, approximate approaches, despite being still expensive, usually fail when the assumptions made in their derivations do not hold. Thus, based on experimental and analytical insights gained from extensive simulations and investigations in different adaptive filtering settings, we propose a new scheme to control the projection order of the APA by monitoring variations of the filter output error, which is considered to be indicative of the adaptation state. This scheme explicitly maximizes \( \Delta(n) \) at each time instance. In what follows, we explain the proposed scheme and its rationale.

In the presence of background noise and after the initial convergence when the variance of the filter output error becomes comparable to the background noise power, a large projection order degrades the performance and increases the steady-state MSE by amplifying the background noise. On the other hand, a small projection order drives the algorithm towards convergence slower at early stages of the adaptation, but achieves a lower steady-state MSE as a result of reduced noise amplification. Hence, in order to have a fast convergence and reach a low steady-state MSE, the projection order should be set as large as possible at the start-up and kept large as long as the variance of the filter output error is considerably higher than the noise power. Then, when the algorithm approaches its steady state and the variance of the error becomes comparable to the noise power, the projection order should be reduced to counter the noise amplification. This leads to smaller steady-state MSE as well as reduced complexity.

It is inferred from the above discussion that to achieve a desirable performance, the projection order needs to be changed depending on the variance of the filter output error. Of course, this dependence should be formulated in a proper way. Recalling that the learning curve of the APA exhibits an exponential decay in time, evidenced by both theoretical analyses (see, e.g., \([13], [17], [18]\) and simulations (see, e.g., Figs. 3 and 4), we propose to define a linear dependence of the projection order on the logarithm of the error variance. In other words, we let the projection order take integer values between 1 and \( k_{\text{max}} \) at each time instant while linearly depending on the logarithm of the error variance. The projection order takes its highest value, \( k_{\text{max}} \), as long as the error variance is higher than a threshold, defined by \( \gamma \), and reduces to 1 when the error variance is equal or lower than the predicted steady-state MSE, \( \eta \). Between these thresholds, the projection order is computed depending on the logarithm of the error variance. We can formulate this rule as

\[
k_p(n) = 1 + \left( k_{\text{max}} - 1 \right) \frac{\ln(\sigma^2_p(n)/\eta)}{\ln(\gamma/\eta)} \tag{2a}
\]

\[
k_p(n) = \min\left\{ k_{\text{max}}, \max\{1, \lceil k_p(n) \rceil\} \right\} \tag{2b}
\]

where \( \sigma^2_p(n) \) is an estimate of the error variance, \( k_p(n) \) is the variable projection order, and \( \lceil x \rceil \) denotes the smallest integer greater than or equal to \( x \). According to \((2)\), the projection order is reduced when the algorithm converges (i.e., \( \sigma^2_p(n) \) decreases) and is increased when the algorithm moves away from the steady state and \( \sigma^2_p(n) \) increases. The latter is usually the case with time-varying systems and/or non-stationary signals. The steady-state MSE of the APA, \( \eta \), can be predicted by approximating the associated equation given in \([13]\) as

\[
\eta = \sigma^2_p \left( 1 + \frac{\mu}{2 - \mu} \text{Tr}(R_k) \frac{k}{\|k(n)\|^2} \right) \geq \sigma^2_p \left( 1 + \frac{\mu k}{2 - \mu} \right)
\]

where \( \text{Tr}(R_k) \) is the trace of the input autocorrelation matrix and \( \sigma^2_p \) is the variance of the background noise. The threshold

\[
\sigma^2_p \left( 1 + \frac{\mu}{2 - \mu} \text{Tr}(R_k) \frac{k}{\|k(n)\|^2} \right) \geq \sigma^2_p \left( 1 + \frac{\mu k}{2 - \mu} \right)
\]
\( \gamma \) determines at what level of \( \sigma^2(n) \) the algorithm starts switching the projection order. In order to attain the fastest convergence, it is beneficial to decrease \( \gamma \) for small values of \( k_{\text{max}} \). On the other hand, with relatively large values of \( k_{\text{max}} \), a large \( \gamma \) results in more computational savings without sacrificing the convergence performance. Accordingly, we compute \( \gamma \) as

\[
\gamma = \eta \left( \frac{\sigma^2}{\eta} \right)^z
\]

where the parameter \( z \) is calculated in such a way that \( \gamma \) gets its maximum value of \( \gamma_{\text{max}} = \sqrt{\eta \sigma^2} \) for \( k_{\text{max}} \geq 8 \) and its minimum value of \( \gamma_{\text{min}} = \eta \) for \( k_{\text{max}} = 1 \). This can be realized by defining \( z \) with a linear dependence on \( k_{\text{max}} \) as

\[
z = \max \{ 0, \min \{ 0.5, \frac{k_{\text{max}} - 2}{12} \} \}.
\]

In (3), \( \sigma^2 \) is the power of the reference signal, \( d(n) \). In fact, \( \sigma^2 \) is the initial value for the error variance estimate, \( \sigma^2(0) \), since we initialize the algorithm to \( w(0) = 0 \) so \( e(0) = d(0) - w^T(0)x(0) = d(0) \). The reference signal power, if not known \textit{a priori}, can be estimated using the first \( k_{\text{max}} \) samples of \( d(n) \) during initial population of the input regression matrix.

Fig. 2 illustrates the proposed rule for controlling the projection order by depicting the time-varying continuous projection order, \( k(n) \), as a linear function of the logarithm of the error variance, \( \ln \sigma^2(n) \), together with the bounded and quantized integer projection order, \( \kappa(n) \).

From a practical point of view, in order to achieve a more reliable and smooth behavior, we can adjust the variable projection order in a manner that it takes only power-of-two values, i.e., 1, 2, 4, 8 ... Therefore, we can replace (2b) with

\[
\kappa(n) = \min \{ k_{\text{max}}, 2^{\lfloor \log_2 \max \{ 1, \kappa(n) \} \rfloor} \}.
\]

This equation resembles a quantization function and can be evaluated using a lookup table. We call the proposed algorithm, which uses \( \kappa(n) \) instead of a fixed \( k \), \textit{affine projection algorithm with variable projection order} (APA-VPO).

V. ERROR VARIANCE ESTIMATION

The most commonly used method for estimating the error variance is the well-known exponential window approach [2], [3]:

\[
\sigma^2(n) = \lambda \sigma^2(n-1) + (1-\lambda)e^2(n)
\]

where \( \lambda \) is a forgetting factor and the initial value is \( \sigma^2(0) = \sigma^2 \). Although this method is widely used, its accuracy is very sensitive to the value of \( \lambda \) and it may fail to yield a reliable estimate in all stages of adaptation mainly because of using a fixed \( \lambda \) [5], [19]. A smaller value of \( \lambda \) allows better tracking during the transient state and a larger value of \( \lambda \) results in smoother estimation close to and at the steady state. Hence, we propose to control \( \lambda \) using a rule similar to (2), viz. to reduce it at the transient state and increase it when approaching the steady state. Therefore, the time-varying forgetting factor is adapted via

\[
\lambda(n) = \alpha_{\text{min}} + (\alpha_{\text{max}} - \alpha_{\text{min}}) \frac{\ln(\sigma^2(n-1)/\sigma^2)}{\ln(\eta/\sigma^2)}
\]

where \( \alpha_{\text{max}} \) and \( \alpha_{\text{min}} \) are maximum and minimum allowed values for \( \alpha(n) \), respectively. Experiments suggest that selecting suitable values for \( \alpha_{\text{max}} \) and \( \alpha_{\text{min}} \) is a more tractable task than choosing an appropriate fixed \( \lambda \) for (4). Furthermore, incorporating a simple moving average filtering scheme with the exponential window approach improves the estimation accuracy appreciably at the expense of a slight increase in computations. Therefore, we modify (4) to

\[
\sigma^2(n) = \alpha(n)\sigma^2(n-1) + (1-\alpha(n)) \cdot \text{SMA}(e^2(n), N_m)
\]

where the simple moving averaging (SMA) is implemented as

\[
\text{SMA}(e^2(n), N_m) = \left( e^2(n) + e^2(n-1) + \cdots + e^2(n-N_m+1) \right) / N_m.
\]

Here, \( N_m \) is the moving average window size that is usually set to \( L/2 \).

VI. SIMULATION RESULTS AND DISCUSSIONS

In this section, we present some simulation results to examine performance of the proposed algorithm and compare it with the most relevant previously proposed algorithms.

A. System Identification

In the first part of the simulations, we examined identification of an unknown system with 32 taps. We generated the taps randomly and normalized the unknown system to unit energy. We also assumed the same number of taps for the corresponding adaptive filters (\( L = 32 \)) and initialized them to zero. The maximum projection order was \( k_{\text{max}} = 16 \), the noise variance \( \sigma^2 = 10^{-3} \), the regularization parameter \( \delta = 10^{-4} \), step-size \( \mu = 0.5 \), minimum and maximum values for the forgetting factor \( \alpha_{\text{min}} = 0.2 \) and \( \alpha_{\text{max}} = 0.99 \), and the moving average window size \( N_m = 16 \). The results were obtained by ensemble-averaging over \( 10^4 \) independent trials.

Fig. 3 plots learning curves of the APA-VPO, the NLMS algorithm (APA with \( k = 1 \)), and the APA with different projection orders (\( k = 2 \) to 16) when the input is a zero-mean unit-variance Gaussian signal. To highlight the time indices and the MSE levels at which the projection order is switched in APA-VPO, its learning curve is shown in different colors for different values of its time-varying projection order, \( \kappa(n) \). It is clear that APA-VPO converges as fast as the APA with the highest projection order (\( k = 16 \)) and reaches the smallest steady-state MSE which is the same as that of the NLMS algorithm. In Fig. 4, APA-VPO is compared with E-APA, DSAPA, and SR-APA proposed in [8], [9], and [12], respectively. For SR-APA, the number of input regressors that were selected out of \( k_{\text{max}} \) regressors was set to \( k_{\text{SR}} = 12 \). Input signal was generated by passing a zero-mean unit-variance Gaussian white noise through an autoregressive moving average process, ARMA(2,2), with transfer function.
In order to compare the capability of the algorithms in tracking sudden system variations, the unknown target system was altered with another randomly generated one, half way through the simulations. The upper part of Fig. 4 shows MSE curves of different algorithms and the lower part depicts the ensemble-averaged time-varying projection order of the variable-projection-order algorithms, i.e., E-APA, DS-APA, and APA-VPO.

B. Network Echo Cancelation

In the second part of the simulations, we considered an application of network echo cancelation. We used the echo path model 4 from ITU-T recommendation G.168 [20] and a real recorded voice signal to compare performance of different algorithms in Fig. 5 in terms of the normalized misalignment, i.e., $|\mathbf{w}_n - \mathbf{w}(n)|^2/|\mathbf{w}_n|^2$. Fig. 6 shows the time-varying projection order of E-APA, DS-APA, and APA-VPO for this experiment. In order to compare tracking capabilities of the algorithms, an abrupt change of the echo path was introduced at the middle of the simulation by shifting the impulse response to the right by 10 samples. Simulated parameters were $L = 128$, $k_{\text{max}} = 32$, $k_{\text{SR}} = 24$, $\sigma_n^2 = 10^{-2}$, $\delta = 10^{-3}$, $\mu = 0.5$, $\alpha_{\text{min}} = 0.2$, $\alpha_{\text{max}} = 0.99$, and $N_m = 64$.

C. Discussion

It is clearly seen in the simulations that APA-VPO outperforms its contenders in terms of both convergence rate and steady state MSE since it tunes the projection order in a better way than E-APA and DS-APA. In particular, APA-VPO reduces to the NLMS algorithm at the steady state whereas the other algorithms lack this desirable property. Reducing to the NLMS algorithm, in addition to decreasing the steady-state misalignment, alleviates the average computational load dramatically since the NLMS algorithm needs no matrix inversion (or solution of any linear system of equations). We may attribute APA-VPO’s superiority to two aspects: first, the underpinning idea of tuning the projection order in a sensible manner by monitoring the adaptation state; second, attaining a good estimate of the error variance, $E[e^2(n)]$, which faithfully represents the adaptation state, rather than relying on the instantaneous value of the squared error, $e^2(n)$, which is the case in E-APA and DS-APA.

VII. CONCLUSION

A new affine projection algorithm was proposed with a variable projection order that is controlled by the estimated error variance. The error variance is estimated using a new method, which takes advantage of both exponential window and simple moving average techniques together with a time-varying forgetting factor. In addition to its superior performance in terms of high initial convergence rate and low steady-state misalignment, the proposed algorithm is able to track system variations effectively. Reduced average computational burden is one of the most significant benefits of the proposed algorithm as it intends to minimize the projection order without sacrificing the convergence performance.

REFERENCES


Fig. 1. MSD gradient of APA in a typical system identification set-up with $L = 16$, $\mu = 0.3$, SNR = 10 dB, and Gaussian input signal.

Fig. 2. Time-varying continuous projection order, $\hat{\kappa}(n)$, as a linear function of the logarithm of the error variance, $\ln s(n)^2$, together with the bounded and quantized integer projection order, $\kappa(n)$.

Fig. 3. MSE curves of APA-VPO, NLMS, and APA with different projection orders ranging from 2 to 16. Input is Gaussian and $\mu = 0.5$.

Fig. 4. MSE and ensemble-averaged time-varying projection order of different algorithms for Gaussian ARMA(2,2) input.

Fig. 5. Normalized misalignment of different algorithms for network echo path identification using voice input.

Fig. 6. Time-varying projection order of different algorithms for the experiment of Fig. 5.