

On Affine Projection Direction Vector (Non-)Whiteness

A. A. (Louis) Beex

DSP Research Laboratory – ECE 0111
Virginia Tech
Blacksburg, VA 24061-0111

Sundar G. Sankaran

Intel
Broadband Wireless Division
Santa Clara, CA 95054

Abstract—Hypotheses made about the behavior of the affine projection algorithm (APA) are investigated by means of carefully designed comparative experiments. These experiments lead us to the properties of the direction vector used in APA. Experiments also show that APA converges faster than NLMS in the case of white noise input, leading us to question that whitening of the direction vector is the explaining factor for the convergence acceleration observed with APA. An alternative hypothesis is offered for why APA may yield faster convergence than NLMS.

I. INTRODUCTION

The affine projection algorithm (APA) has been advocated as an answer to the slow convergence of the normalized least mean square algorithm (NLMS) when its input is highly correlated. When the input is an autoregressive process, and the APA order is higher than the order of that AR process, convergence is accelerated dramatically. The direction vector used by APA, in this AR input case, was hypothesized to be white. The whiteness of the direction vector was hypothesized to be the cause of the accelerated convergence [1]. The latter hypotheses have been used to advance analyses of APA [2, 3].

While under many experimental conditions the above hypotheses are approximated quite well by actual behavior, careful experiments show that – strictly speaking – neither of the above hypotheses is valid. A new hypothesis is offered for the cause of convergence acceleration in APA.

The paper is organized as follows. Section II provides the update equation of the APA class of algorithms, as well as the alternative algorithms used for comparison. The experimental setup is explained in Section III. Experimental results are shown and discussed in Section IV and, towards our alternative hypothesis, in Section V.

II. THE APA ALGORITHM

Figure 1 shows an adaptive filter used in the system identification mode. The wide sense stationary input process x_n to the system $H(z)$ and the corresponding measured output d_n , possibly contaminated with measurement noise v_n , are known. The unit variance, zero-mean, white noise input process ε_n is colored by the filter $H_x(z)$ resulting in the (colored) input process x_n . The input process is converted into input vectors \mathbf{x}_n , via a tapped delay line (TDL), which are then used by the adaptive algorithm.

The input vector \mathbf{x}_n is defined as

$$\mathbf{x}_n = [x_n \quad x_{n-1} \quad \cdots \quad x_{n-M+1}]^T. \quad (1)$$

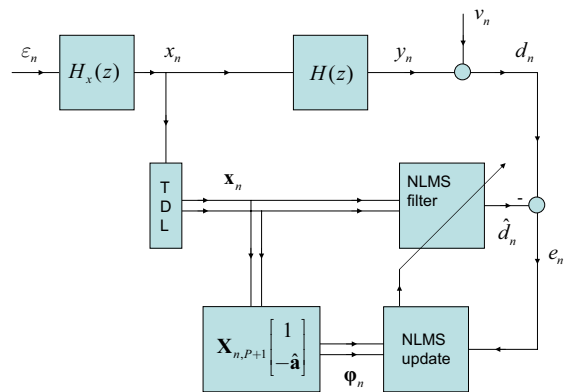


Fig. 1: System identification test set-up.

The objective is to estimate an M -dimensional weight vector \mathbf{w}_n such that the estimated output from the adaptive filter \hat{d}_n is as close as possible to the measured output d_n in mean-squared error sense. The adaptive filter implements an NLMS-like update of the weight vector, as follows:

$$\begin{aligned} \hat{d}_n &= \mathbf{w}_n^H \mathbf{x}_n \\ e_n &= d_n - \hat{d}_n \end{aligned} \quad (2)$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \frac{e_n^*}{\Phi_n^H \Phi_n} \Phi_n$$

where the direction vector Φ_n is evaluated according to

$$\begin{aligned} \Phi_n &= \mathbf{x}_n - \mathbf{X}_{n-1,P} \hat{\mathbf{a}}_n \\ &= \mathbf{X}_{n,P+1} \begin{bmatrix} 1 \\ -\hat{\mathbf{a}}_n \end{bmatrix} \end{aligned} \quad (3a)$$

with $\mathbf{X}_{n-1,P}$ and $\mathbf{X}_{n,P+1}$ defined from

$$\mathbf{X}_{n-n_0,P} = [\mathbf{x}_{n-n_0} \quad \mathbf{x}_{n-n_0-1} \quad \cdots \quad \mathbf{x}_{n-n_0-(P-1)}] \quad (3b)$$

Note that the first subscript of $\mathbf{X}_{p,q}$ defines its leftmost column \mathbf{x}_p according to (1), while the second subscript defines the total number of such columns decrementing the index of the leftmost column. The above formulation allows us to implement several algorithms of interest.

We are interested specifically in the behavior of APA, when it is used to implement the adaptive filter. APA is characterized by the use of $\hat{\mathbf{a}}_n$, evaluated by solving the following equations in least-squares sense,

$$\mathbf{X}_{n-1,P} \hat{\mathbf{a}}_n = \mathbf{x}_n \quad (4a)$$

or equivalently

$$\left[\mathbf{X}_{n-1,P}^H \mathbf{X}_{n-1,P} \right] \hat{\mathbf{a}}_n = \mathbf{X}_{n-1,P}^H \mathbf{x}_n \quad (4b)$$

The algorithm in (2) - (4) is referred to as APA(P) [4]. The constant μ is usually referred to as the step-size. For a wide sense stationary environment, as the step-size vanishes, algorithm performance approaches that of the corresponding time-invariant Wiener filter. However, in some applications fast convergence is of primary interest and fastest convergence is obtained when the step-size equals one. Unless specifically noted otherwise, the step-size will be assumed to equal one in the experiments that follow. Note that $\hat{\mathbf{a}}_n$ in (4) is obtained from the orthogonal projection of \mathbf{x}_n onto the span of $\mathbf{X}_{n-1,P}$, i.e. from the orthogonal projection of the present input vector onto the most recent P (past) input vectors. Consequently, the direction vector $\boldsymbol{\varphi}_n = \mathbf{x}_n - \mathbf{X}_{n-1,P} \hat{\mathbf{a}}_n$ is the error vector in estimating (in least squares sense) the present input vector from the P most recent input vectors. By the orthogonality principle the direction-vector $\boldsymbol{\varphi}_n$ is orthogonal to the P most recent input vectors, i.e. $\boldsymbol{\varphi}_n \perp \mathbf{X}_{n-1,P}$. Note also that $\boldsymbol{\varphi}_n \in sp(\mathbf{X}_{n,P+1})$.

We will be comparing APA(P) to NLMS, which is seen to result from replacing $\hat{\mathbf{a}}_n$ in (3) by $\hat{\mathbf{a}}_n \equiv \mathbf{0}$. The latter can be interpreted as having projected onto an empty set of most recent input vectors, i.e. $P=0$. Therefore, NLMS is equivalent to APA(0) and uses the direction vector $\boldsymbol{\varphi}_n = \mathbf{x}_n$. We observe that the APA(P) algorithm is based on instantaneous quantities, i.e. there are no statistical characterizations involved.

The initialization of APA(P) makes a difference. We will use APA(P) to refer to initialization that begins with an NLMS iteration (i.e. APA(0)), followed by APA(1), followed by APA(2), etc., until an APA(P) iteration is reached, and then continually repeated. Soft initialization of $\mathbf{X}_{n-1,P}$ with P data vectors and executing APA(P) iterations at each iteration will be referred to as APAX(P), and may provide a better indicator of performance in a non-stationary environment.

Note that $\hat{\mathbf{a}}_n$ can also be used to provide the statistically optimal estimate of the input vector \mathbf{x}_n based on $\mathbf{X}_{n-1,P}$, leading to the Wiener-Hopf equation.

$$E \left[\mathbf{X}_{n-1,P}^H \mathbf{X}_{n-1,P} \right] \mathbf{a} = E \left[\mathbf{X}_{n-1,P}^H \mathbf{x}_n \right] \quad (5)$$

In the latter case, $\hat{\mathbf{a}}_n$ equals the order $P-1$ predictor polynomial, $\hat{\mathbf{a}}_n = \mathbf{a}$, which is no longer a function of n . Note that in this case there is no instantaneous orthogonal projection; instead the orthogonal projection is now one defined in the statistical sense. For the prediction error case we have $\boldsymbol{\varphi}_n = \boldsymbol{\varepsilon}_n$, where $\boldsymbol{\varepsilon}_n$ is a tapped delay line vector of prediction errors. For sufficiently high order $P-1$, the latter will be innovations.

III. EXPERIMENTAL SETUP

For ease of interpretation of the results, the system $H_x(z)$ is normalized so that its output is unit variance. After that, $H(z)$ is also normalized so that its output is unit variance. Consequently, the smallest mean square error we expect to possibly see is the variance σ_v^2 of the white measurement noise process ν_n , while the largest mean square error is expected to be one (worst-case, but optimal).

For a wide sense stationary, zero-mean, white noise process, any unitary basis can serve as the eigen-decomposition of the correlation matrix.

$$\mathbf{R}_{xx} = E \left[\mathbf{x}_n \mathbf{x}_n^H \right] = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^H \quad (6)$$

For our comparative purposes, this means that the unitary eigen-decomposition matrix for an arbitrary process can be used to determine whether a process is white. Whiteness implies that all eigen-values are equal. The energy mass distribution of \mathbf{x}_n , over its various eigen-directions, is given by the ratio of the eigen-values to the trace.

$$p_i = \frac{\lambda_i(\mathbf{R}_{xx})}{tr(\mathbf{R}_{xx})} \quad (7)$$

For a white process, the energy distribution is uniform across all eigen-directions. We will use the unitary eigen-matrix \mathbf{V} corresponding to the input process $\{\mathbf{x}_n\}$ for one of the subsequent comparisons.

One measure of whiteness is the constancy of the energy distribution, or the set of eigen-values. The corresponding indicator of rotational (spherical) invariance ρ is defined as

$$\rho_\varphi = \frac{M \left(\prod_{i=1}^M p_i \right)^{-M}}{\sum_{i=1}^M p_i} \quad (8)$$

Another interpretation of the whiteness of the direction vector is to consider whether its correlation matrix is a diagonal matrix, but with elements that are not necessarily equal. The following measure of diagonal dominance,

$$\delta_\varphi = 1 - \frac{\left\| \mathbf{R}_\varphi - \text{diag} \left\{ \text{diag}(\mathbf{R}_\varphi) \right\} \right\|_F}{\left\| \mathbf{R}_\varphi \right\|_F} \quad (9)$$

is therefore an indicator of whiteness. The Frobenius norm is used and the correlation matrix is estimated from averaged direction vector outer products.

The weight vector \mathbf{w}_* , of the system to be identified, is chosen to be the maximum entropy vector, i.e. equal weight is given to each of the eigen-directions.

$$\mathbf{w}_* = \mathbf{V}\mathbf{1} \quad (10)$$

As mentioned, the system weight vector is scaled so that the system output is unit variance.

The various algorithms are initialized with an all-zero weight vector, and – for APAx – with the past data, i.e. $\mathbf{X}_{n-1,P}$, loaded.

IV. EXPERIMENTAL RESULTS

The number of iterations in each experiment is 1,000. Each experiment is repeated 1,000 times. The mean square error is the average of the instantaneous squared errors. The order of the system and the model are both 9 ($M=10$). The output signal-to-measurement noise ratio is set to 30 dB.

4.1 Experiment 1 - The input process is complex, white, Gaussian noise.

Fig. 2 shows the MSE learning curves for the APA(3), APAx(3), APA(7), APAx(7), and NLMS algorithms.

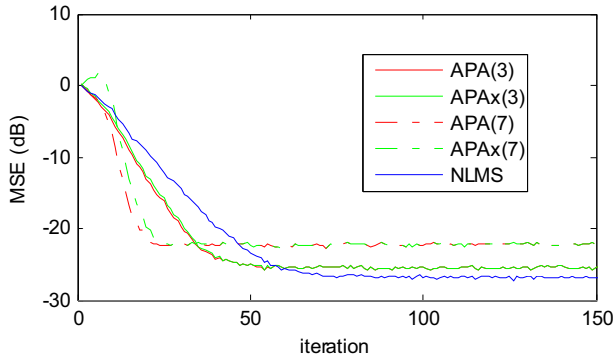


Fig. 2: MSE for APA(3), APA(7), and NLMS, for white input.

We observe that APA(3) converges about 30% faster than NLMS, as it reaches its steady-state value in about 60 rather than 90 iterations. As expected, the steady-state MSE value for APA(3) is higher than for NLMS. There is little difference in the behavior of APA(3) and APAx(3). APA(7) converges in about 25 iterations; more than two times as fast as APA(3) and three times as fast as NLMS. Again, we have a larger misadjustment. APAx(7) takes a little longer to get going, but then converges at the same rate and to the same MSE value as APA(7). The additional subspace dimension increases the convergence rate, even though the input process is a white process.

Fig. 3 shows the relative energy distribution (RED) of the direction vectors for APA(3) and NLMS, determined by projection onto the eigenvectors of the system input process x_n , in addition to its theoretical expectation according to (7).

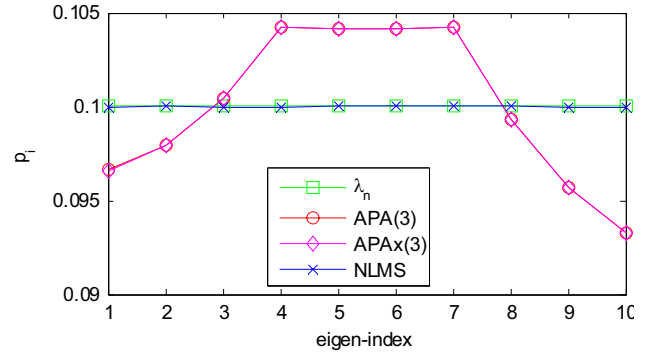


Fig. 3: Direction vector RED for APA(3) and NLMS for white input.

The NLMS direction vectors are white, as expected in theory for a white process. As seen in Fig. 3, the APA direction vectors are “not quite white.” The spherical invariance of the relative energy distribution, as measured according to (8), is 0.9992, 0.9992, and 1.000 for APA(3), APAx(3), and NLMS resp. Note that the spherical invariance measure is insensitive relative to the visual non-constant behavior for APA. Computing the spherical invariance based on the eigenvalues of the estimated direction vector correlation matrix yields 0.9992, 0.9992, and 1.0000 for APA, APAx, and NLMS, respectively, i.e. these values correspond directly to those based on the RED. The estimated direction vector correlation matrix is also used to compute the diagonal dominance, which yields 0.9981, 0.9981, and 0.9964 for APA(3), APAx(3), and NLMS, respectively.

Fig. 4 shows the relative energy distribution of the direction vectors for APA(7) and NLMS, together with the theoretical expectation based on the input process.

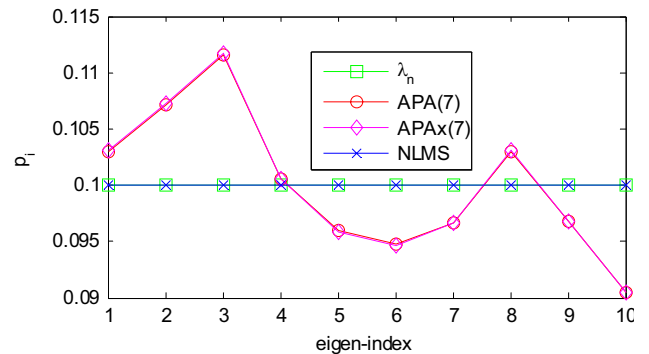


Fig. 4: Direction vector RED for APA(7) and NLMS for white input.

The NLMS direction vectors are white, while the APA(7) direction vectors are not. In fact, the APA(7) direction vectors deviate from white more than the APA(3) direction vectors did. The latter observation is visual, in terms of a loss of flatness, as well as indicated by the RED based numerical values for spherical invariance of 0.9982 and 0.9982, for APA(7) and APAx(7), respectively. Based on the direction vector correlation matrix estimate we find the same spherical invariance values as above and diagonal

dominance values of 0.9976 and 0.9976, respectively, for APA(7) and APAx(7).

For a white input process to the system, APA(7) yields faster convergence than APA(3), which in turn yields faster convergence than NLMS. Simultaneously the APA(7) direction vectors are less white than the APA(3) direction vectors, which themselves are less white than the NLMS direction vectors. We conclude that it is not whiteness that is responsible for faster convergence.

4.2 Experiment 2 - The input is now changed to a colored process. The AR(1) input process is determined by its pole, $p=0.95\exp(2\pi j0.2)$, corresponding to the ideal predictor polynomial $a=[-0.2936-0.9035j]$. The AR(1) system is driven by zero-mean complex, white, Gaussian noise of unit variance. Recall that the systems in Fig. 1 are scaled so that their respective outputs are unit variance.

Figure 5 shows the MSE learning curves for APA(3) and NLMS when the input process is AR(1).

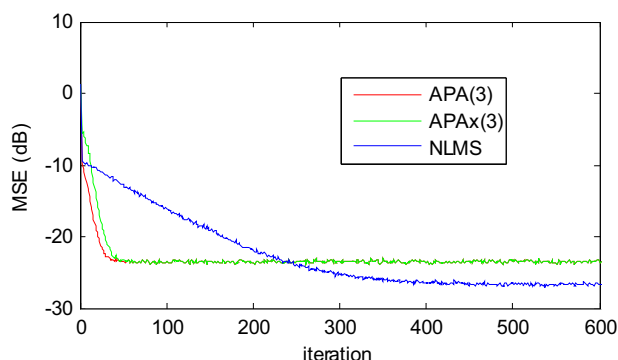


Fig. 5: MSE for APA(3) and NLMS, for AR(1) input.

The convergence of NLMS has slowed considerably, in comparison with the white noise case shown in Fig. 2, while the convergence rate of APA(3) has remained approximately the same. APA and NLMS show fast initial convergence, as they adjust in the major eigen-direction first, while APAx does not.

Fig. 6 shows the relative energy distribution for the direction vectors for APA(3), APAx(3), and NLMS together with what is expected in theory, according to (7), for the AR(1) input process.

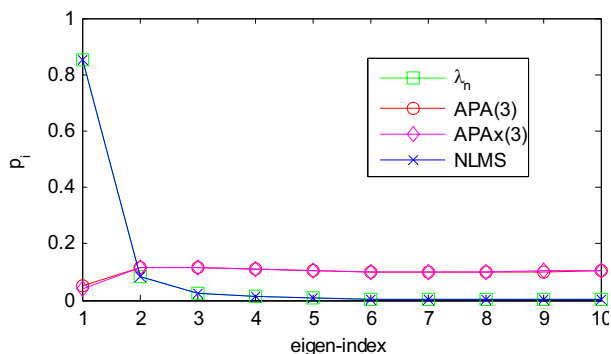


Fig. 6: Direction vector RED for APA(3) and NLMS, for AR(1) input.

The APA(3) direction vectors are “whiter” than those for NLMS, but considerably less white – visually as well as numerically, with spherical invariance values of 0.9807, 0.9680, and 0.1227 for APA, APAx, and NLMS resp.– than previously for the white input. However, neither set of direction vectors can be considered white. The correlation matrix estimate based spherical invariance indicators were 0.9799, 0.9672, and 0.1227, while the diagonal dominance indicators were 0.8294, 0.7973, and 0.0706, for APA, APAx, and NLMS, respectively. The latter again give a good indication of direction vector non-whiteness.

4.3 Experiment 3 - To look at the effect of increased correlation, or coloration, we use a narrowband first-order autoregressive process, denoted NBAR(1), by moving the pole closer to the unit circle, $p=0.99\exp(2\pi j0.2)$. The predictor polynomial is now $a=[-0.3059-0.9415j]$.

Fig. 7 shows the MSE learning curves for APA(3), APA(7), and NLMS, for the NBAR(1) input process.

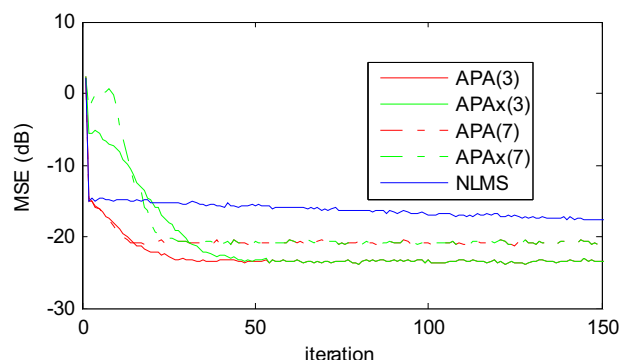


Fig. 7: MSE for APA(3), APA(7), and NLMS, for NBAR(1) input.

The convergence rate of APA(3) seems not to have been affected, while the convergence of NLMS has slowed again (it takes 1,000 iterations to reach steady-state) relative to when the AR(1) input was used (Fig. 5). APA(7) converges faster to its steady-state MSE than APA(3), but does so at a similar rate. The APAx take a little longer to converge than the corresponding APA, due to their different initialization.

Figure 8 shows the direction vector RED for APA(3) and NLMS for the NBAR(1) input process, together with the theoretical expectation for the latter input process.

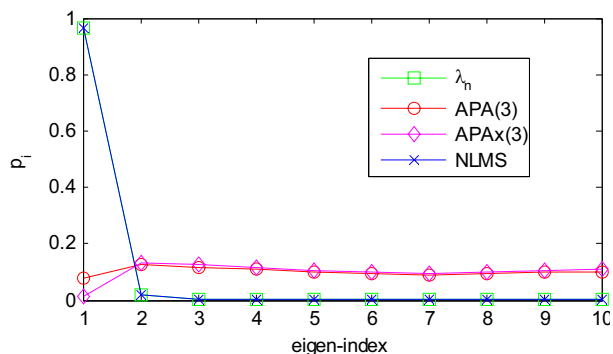


Fig. 8: Direction vector RED for APA(3) and NLMS, for NBAR(1) input.

While the APA(3) direction vectors are not white, they are “whiter” than those for NLMS. As the correlation has increased, relative to the AR(1) process, the APA(3) direction vectors have become more white – as indicated also by the spherical invariance indicators of 0.9914, 0.8952, and 0.0296 for APA, APAx, and NLMS resp. – relative to the previous case shown in Fig. 6. Note that the APA(3) direction vectors have become more white, and that the convergence rate – relative to NLMS – has increased. The estimated direction vector correlation matrix yields spherical invariance indicators of 0.9904, 0.8942, and 0.0296, and diagonal dominance values of 0.8617, 0.7072, and 0.0549, for APA, APAx, and NLMS, respectively, which corresponds well with the visual behavior in Fig. 8.

Note however, that all these indicators point to the APA direction vectors not being white. The discrepancy from whiteness is largest at eigen-index 1, i.e. at the major eigenvalue. The discrepancy is more severe for APAx, which uses APA(P) iterations all the time, even at startup, implying that almost no weight adjustments take place in the direction of the major eigenvector. Since APA starts with an NLMS iteration, its first weight adjustment is generally in the direction of the major eigenvector. Afterwards there is not much energy in the direction vectors to make further adjustments in the major eigen-direction.

There is a major difference in spherical invariance between APA and APAx, which is due to the difference in their initializations.

Figure 9 shows the direction vector RED for APA(7) and NLMS for the NBAR(1) input process, together with the theoretical expectation for the latter input process.

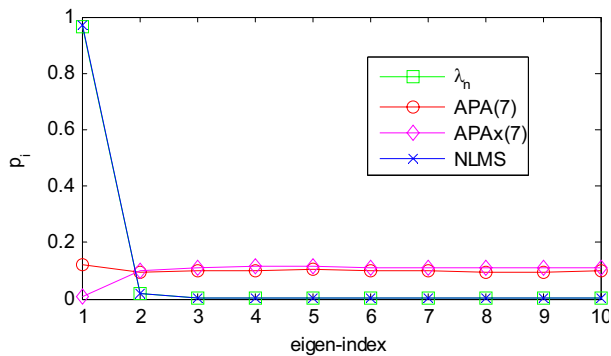


Fig. 9: Direction vector RED for APA(7) and NLMS, for NBAR(1) input.

The RED based spherical invariance values are 0.9978, 0.8321, and 0.0293, for APA, APAx, and NLMS, resp. While spherical invariance for APA(7) has increased, relative to APA(3), the steady state adjustments occur even less in the direction of the major eigenvector. This is indicated by the diminishing spherical invariance parameter for APAx, as well as by the reduction of energy at eigen-index 1 in Fig. 9 relative to Fig. 8. The direction vector correlation matrix estimate yields spherical invariance indicators of 0.9953, 0.8321, and 0.0293, and diagonal dominance indicators of 0.9098, 0.6957, and 0.0548, for APA, APAx, and NLMS, respectively, supporting the

observations made above.

Increasing the correlation of the input process – and thereby the eigenvalue ratio – leads to the major eigenspace becoming more and more dominant. Consequently, the span of $\mathbf{X}_{n-1,P}$ tends to be the span of the major eigenspace. As the direction vector of APA is orthogonal to $\mathbf{X}_{n-1,P}$, the directions of adjustment tend not to be in the major eigenspace. This is observed in Figs. 6 and 8, where eigen-index 1 shows that there is little energy in the direction of the major eigen-space once the APA iterations are under way. At the same time, the input to the system tends to lie in the major eigenspace. The latter excitation yields a desired signal determined by the major eigendirection, while the corrections of the weight vector are occurring in the directions that tend not to lie in the major eigenspace. Note in Figs. 5 and 7 that the initial weight vector corrections in NLMS and APA, which tend to be in the direction of the major eigenspace, are much more effective in reducing MSE than the initial weight vector corrections of APAx, which tend to be in a direction orthogonal to the major eigenspace. These tendencies are stronger as the input process is more predictable from its own past, i.e. as the innovations variance of the AR process is relatively smaller.

4.4 Experiment 4 - As the direction vectors in APA correspond to an instantaneous least squares prediction error vector, one might think that replacing that direction vector with its ideal counter-part would provide the limiting performance possible with APA. To test this hypothesis we now return to using the AR(1) input process of Experiment 1 and let $\hat{\mathbf{a}}_n$ equal the order $P-1$ predictor polynomial, $\hat{\mathbf{a}}_n = \mathbf{a}$. The latter is now constant, depending only on the statistical characterization of the input process as indicated in (5). This ideal predictor based modification of the APA(3) algorithm is indicated by the acronym APAp(3), for the sake of simplicity, even though the affine projection interpretation no longer holds.

Fig. 10 shows the RED for the direction vectors used in APAp(3), APAxp(3), and NLMS.

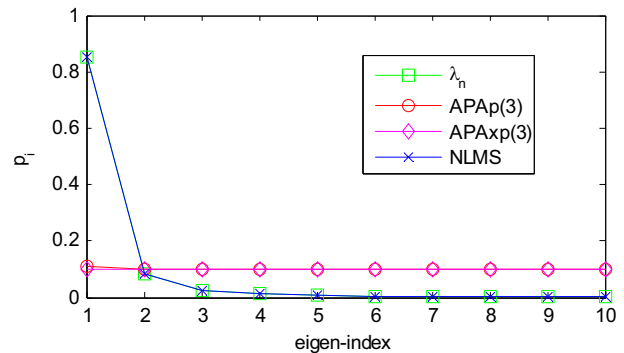


Fig. 10: Direction vector RED for APAp(3) and NLMS, for AR(1) input.

The direction vector for APAp(3) is almost perfectly white. The minor discrepancy is a result of the APA startup

procedure. The direction vector for APAXp(3) has been rendered perfectly white because it starts with a loaded data matrix. The limiting result, of a white prediction error, engenders the MSE learning curves shown in Fig. 11.

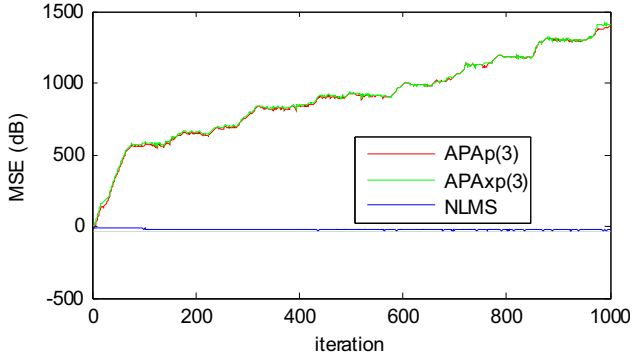


Fig. 11: MSE for APAp(3) and NLMS, for AR(1) input.

Observe that APAp(3) diverges, in spite of its white direction vector. The cause for this divergence is that the direction vector amplitude is determined by the white noise driving process rather than being directly related to the energy in the input process, which is determined by the pole location of the AR(1) process, and which relates to the error signal driving the weight update.

4.5 Experiment 5 - The experiment reflected in Fig. 12 is for an AR(2) process, with poles $p_1=0.95\exp(2\pi j0.2)$ and $p_2=0.9\exp(2\pi j0.4)$. The corresponding predictor polynomial is $\mathbf{a}=[0.4345-0.3745j \ 0.2642-0.8132j]$.

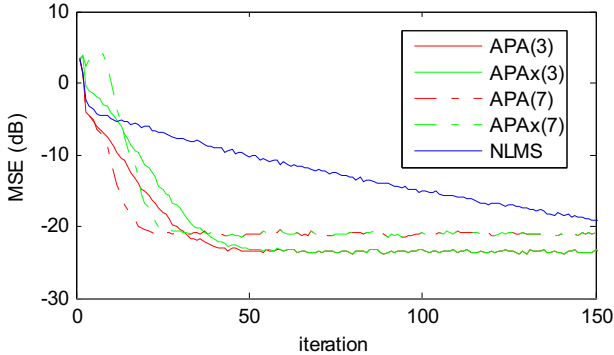


Fig. 12: MSE for APA(3) and NLMS, for AR(2) input.

Note that the convergence rates for APA(3) and APA(7) are similar to those observed earlier (Figs. 2, 5, and 7) for the white and AR(1) processes, and apparently not affected much by the increase in the AR process order.

However, the direction vector energy distribution for APA(3), in Fig. 13, is different from that observed in Figs. 3, 6, and 8. The spherical invariance indicators based on the estimated correlation matrix for the direction vectors are 0.9715, 0.9655, and 0.2850, respectively, while the diagonal dominance indicators are 0.7962, 0.7805, and 0.1345, respectively. Again, these indicators point to direction vector non-whiteness.

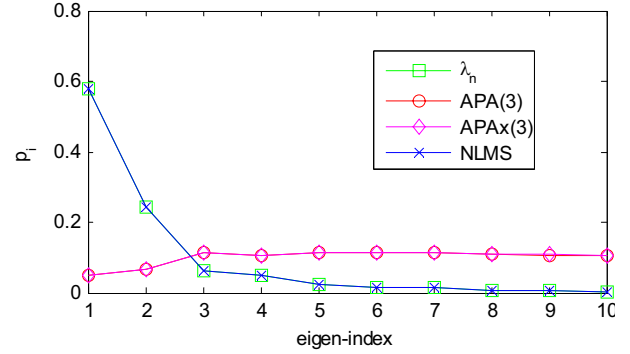


Fig. 13: Direction vector RED for APA(3) and NLMS, for AR(2) input.

The direction vector energy distribution for APA(7), in Fig. 14 is somewhat similar to that in Fig. 13. The estimated direction vector correlation matrix produces spherical invariance indicators of 0.9248, 0.8938, and 0.2849, and diagonal dominance values of 0.7012, 0.6675, and 0.1342, resp. confirming loss of whiteness in the APA(7) direction vectors relative to those for APA(3). Again, we observe a loss of whiteness simultaneous with faster convergence.

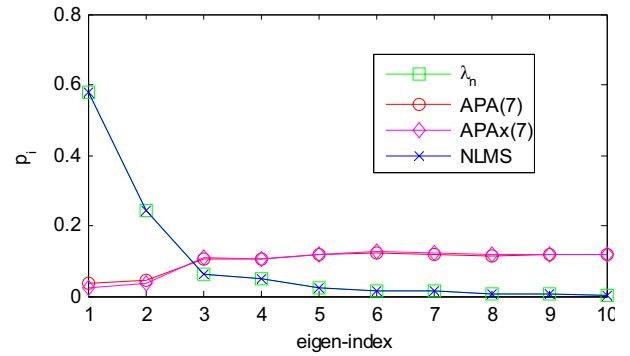


Fig. 14: Direction vector RED for APA(7) and NLMS, for AR(2) input.

V. ALTERNATIVE HYPOTHESIS

So far we looked at whiteness related to the elements of the direction vectors themselves. Our next whiteness test relates one direction vector to another, i.e. they occur at different times. Thereto the normalized absolute correlation (NAC) between direction vectors at lag l is defined as:

$$\gamma = \frac{|\Phi_n^H \Phi_{n-l}|}{\|\Phi_n\| \|\Phi_{n-l}\|}. \quad (11)$$

In addition we look at the NAC between the direction vector and the weight error vector, using this replacement in (11)

$$\Phi_{n-l} \leftarrow \mathbf{w}_n - \mathbf{w}^*. \quad (12)$$

Figures 15 and 16 show the direction vector NAC CDF corresponding to Experiment 1 in 4.1, for white noise input, for APA(3) and APA(7), respectively. The result for NLMS ($P=0$) looks like (it is not shown explicitly) the results for lags 4 and higher in Fig. 15.

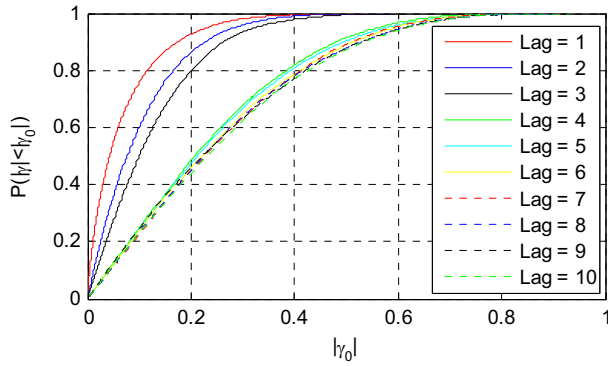


Fig. 15: NAC CDF for APA(3), for white noise input.

We observe that APA(3) produces direction vectors (for lags 1-3) that are less correlated (CDF towards the top-left of the graph) than the direction vectors for NLMS.

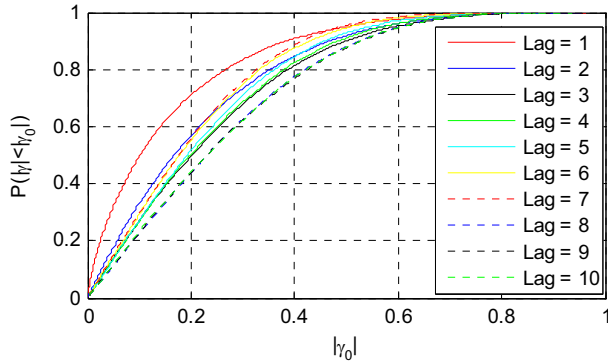


Fig. 16: NAC CDF for APA(7), for white noise input.

Fig. 16 shows that the APA(7) direction vectors are not as uncorrelated as those for APA(3), which corresponds to our earlier observations of not being as white. However, APA(7) converged faster than APA(3), so that whiteness can not be the sole explaining factor for faster convergence.

In Fig. 17 the NAC CDF of the direction and weight error vectors is shown for APA(0) (NLMS), APA(3), and APA(7), for the white noise input case.

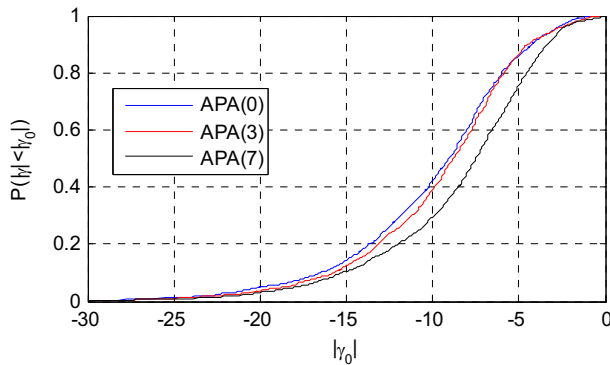


Fig. 17: NAC CDF for weight error, for white noise input.

Here we observe that for APA(3) the direction vector is slightly more correlated with the weight error vector, while

for APA(7) that correlation is stronger. The latter suggests that these direction vectors are more effective in reducing the weight error, resulting in faster convergence.

Figure 18 shows the CDF for the direction vector - weight error vector NAC for the AR(1) input of Example 2 in 4.2.

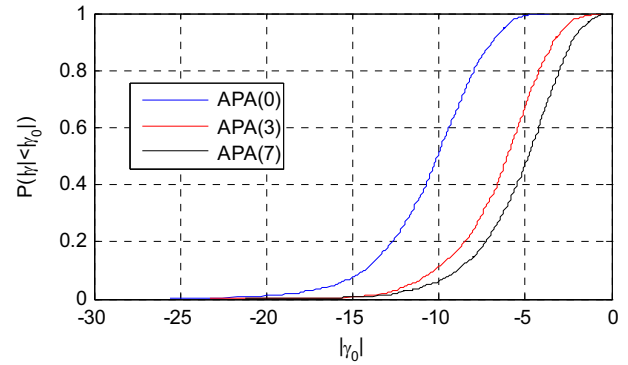


Fig. 18: NAC CDF for weight error, for AR(1) input.

This result shows again that for increasing order APA the direction vector correlation with the weight error vector is progressively stronger. These direction vectors are more effective in reducing the weight vector error, thus producing faster convergence.

For the NBAR(1) input of Experiment 3 in 4.3 the direction vector NAC CDF for NLMS is given in Fig. 19, showing strong correlation.

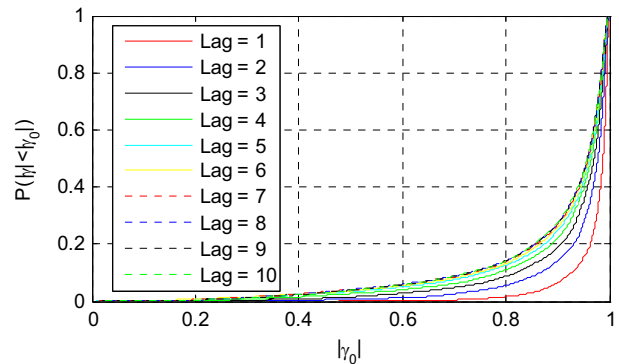


Fig. 19: NAC CDF for NLMS, for NBAR(1) input.

Fig. 20 shows the result of de-correlation by APA(3).

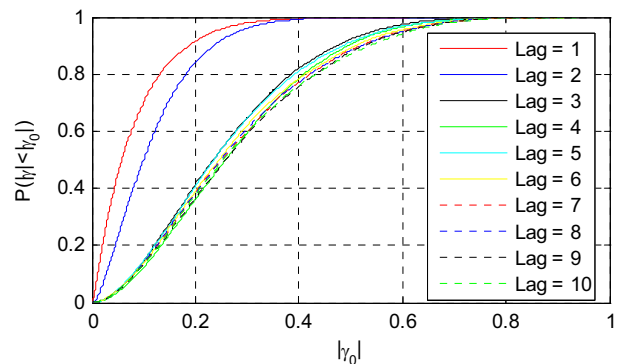


Fig. 20: NAC CDF for APA(3), for NBAR(1) input.

However, Fig. 21 shows that the APA(7) direction vectors are not as de-correlated as those for APA(3), even though the latter converges more slowly.

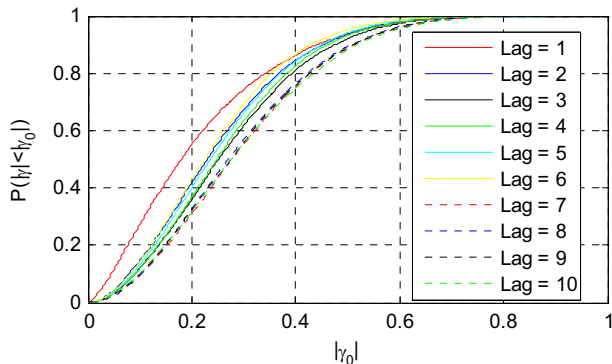


Fig. 21: NAC CDF for APA(7), for NBAR(1) input.

A look at the direction vector – weight error vector correlation, in Fig. 22, shows that for larger APA order the direction vectors are more and more correlated with the weight vector error. This increased correlation with the weight error produces faster convergence.

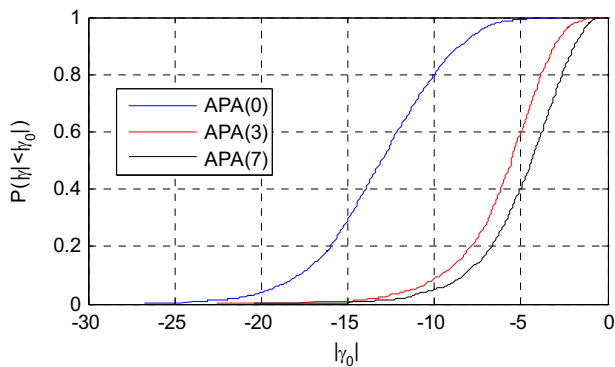


Fig. 22: NAC CDF for weight error, for NBAR(1) input.

In view of the latter experiments, we hypothesize that APA produces direction vectors that have a stronger correlation with the weight error vector for an identified system. As a consequence of that behavior the weight error vector is reduced faster as the order of APA is higher, which is expressed as faster convergence.

The weight error vector, at time instant n , lies in $\mathbf{X}_{n,P+1}^\perp$, i.e. in the subspace orthogonal to all the input vectors then used for adaptation (under the assumption of unit step-size and ignoring measurement noise effects). At time instant $n+1$, APA performs adaptation in $\mathbf{X}_{n,P}^\perp$, i.e. in the subspace orthogonal to the span of $\mathbf{X}_{n,P}$. Thus, the weight error vector and the direction vector live in “almost” the same space. The dimension of this space gets smaller and smaller as P increases, resulting probabilistically in higher and higher correlation between weight error and direction vectors. This leads to faster convergence.

VI. SUMMARY

Examples show that the convergence of APA(P), even for a white noise input process, increases as APA uses a higher dimensional space for its projections, i.e. as P is increased. Also, the direction vectors used in APA – for AR processes of sufficiently low order – are not white. In fact, they can be less white as P is increased, while convergence is faster. Consequently, whiteness of the direction vectors can not be the sole explaining factor for the better convergence properties of APA(P) relative to NLMS. Our hypothesis is that the increase in convergence rate is the result of an increase in correlation between the direction vector and the weight error vector for an identified system. The latter increases as the APA order increases.

VII. REFERENCES

- [1] M. Rupp, "A Family of Adaptive Filter Algorithms with Decorrelating Properties," *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 771 – 775, March 1998.
- [2] N. J. Bershad, D. Linebarger, and S. McLaughlin, "A Stochastic Analysis of the Affine Projection Algorithm for Gaussian Autoregressive Inputs," *ICASSP 2001*, pp. 3837-3840.
- [3] S. J. M. de Almeida, J. C. M. Bermudez, N. J. Bershad, and M. H. Costa, "A Stochastic Model for the Convergence Behavior of the Affine Projection Algorithm for Gaussian Inputs," *ICASSP 2003*, pp. VI-313-316.
- [4] K. Ozeki and T. Umeda, "An Adaptive Filtering Algorithm Using an Orthogonal Projection to an Affine Subspace and Its Properties," *Electronics and Communications in Japan*, Vol. 67-A, No. 5, pp. 19-27, 1984.