An affine projection algorithm with variable step size and projection order

Alberto Gonzalez, Miguel Ferrer, Maria de Diego and Gema Piñero

Institute of Telecommunications and Multimedia Applications (iTEAM), Universitat Politècnica de València
Camino de Vera s/n. 46022-Valencia. SPAIN.
email:{agonzal, mferrer, mdediego, gpinyero}@dcom.upv.es

Abstract

It is known that the performance of adaptive algorithms are constrained by their computational cost. Thus, affine projection adaptive algorithms achieve higher convergence speed when the projection order increases, which is at the expense of a higher computational cost. However, regardless of computational cost, a high projection order also leads to higher final error at steady state. For this reason it seems advisable to reduce the computational cost of the algorithm when high convergence speed is not needed (steady state) and to maintain or increase this cost only when the algorithm is in transient state to encourage rapid transit to the permanent regime. The adaptive order affine projection algorithm presented here addresses this subject. This algorithm adapts its projection order and step size depending on its convergence state by simple and meaningful rules. Thus it achieves good convergence behavior at every convergence state and very low computational cost at steady state.

Keywords: Adaptive filter, Affine projection algorithm, Variable step size, Computational cost.

1. Introduction

Generally speaking affine projection algorithm (APA) [1] is a versatile adaptive strategy that improves the convergence speed of the well known Least Mean Squares (LMS) algorithm, while maintaining good robustness, simplicity and stability. The convergence speed of APA algorithm increases when an integer parameter called projection order (N) is also increased. However its computational cost and final residual error (final misadjustment) get worse at the same time. Variable step-size affine projection algorithms have already been proposed [2] [3] [4] [5] to overcome this duality and achieve better performance in steady state without penalizing the speed of adaptation of the algorithm. Although these strategies achieve better final error in steady state, their computational cost remains invariant throughout algorithm execution and depends mainly on its projection order. Moreover, even when the final residual error is improved, this residual error is not optimum since, as shown in [6], the final steady state error increases with the projection order. Although it should be noted that it can also be shown that the final error can be considered independent of projection order for very small values of the algorithm step-size μ. It seems clear that to minimize the final error at steady state, both the step size of the algorithm and the projection order must be decreased. Moreover the main advantage of reducing the projection order when the algorithm reaches the steady state, provided residual error is decreased, is the reduction in computational cost. Therefore, it appears that the next step of the variable step-size APA evolution is the adaptation of the projection order in response to algorithm performance. A first approach

2

to this idea was proposed in [7], which describes how to combine two affine projection algorithms of different orders such that the result obtained by the combination was better than that obtained for each of the algorithms running independently. Nevertheless, although satisfactory results can be obtained in terms of convergence speed and final residual error, the computational cost of such parallel combination of algorithms can increase considerably, since both algorithms have to work simultaneously. In the present paper an algorithm capable of changing the projection order subject to a rule dependent on algorithm performance is introduced. Thus, when the algorithm reaches its steady state, the projection order is reduced providing a dual positive effect: further minimization of the residual error and decrease in computational cost. This idea of evolving the projection order to improve the affine projection algorithm performance was proposed in [8] and more recently in [9]. However these algorithms adapt the projection order using the instantaneous squared error and keeps a single step-size μ. Even though this lead to improvement due to the change in projection order, as a general rule, it does not achieve optimum residual error in the steady state since this residual error depends on both μ and the projection order. In contrast, the algorithm proposed in the present paper simultaneously adapts the projection order and the step size in a meaningful way, thus it can get profit from both strategies (an alternative technique to this is proposed in [10]). Finally, a few other strategies such as set-membership APA [11] also reduce the computational cost of the APA at steady state because it does not update the coefficients of the adaptive filters at all iterations. This strategy even slightly decreases the final residual error (because the variance of the filter weights is also reduced),

3

but it cannot achieve the overall performance improvements of the algorithm proposed here.

2. Algorithm description

The APA algorithm with regularization computes the filter coefficient vector at each iteration with the following update equation

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{A}^T(n)[\mathbf{A}(n)\mathbf{A}^T(n) + \delta\mathbf{I}]^{-1}\mathbf{e}(n), \qquad (1)$$

that can be rewritten as

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\epsilon(n), \qquad (2)$$

where the regularization factor is denoted by δ, $\mathbf{I}$ is the $N \times N$ identity matrix,

$$\mathbf{A}(n) = \begin{bmatrix} \mathbf{x}^T(n) \\ \mathbf{x}^T(n-1) \\ \vdots \\ \mathbf{x}^T(n-N+1) \end{bmatrix} \quad \mathbf{d}(n) = \begin{bmatrix} d(n) \\ d(n-1) \\ \vdots \\ d(n-N+1) \end{bmatrix}, \quad \text{and } \mathbf{e}(n) = $$
$$\mathbf{d}(n) - \mathbf{A}(n)\mathbf{w}(n-1).$$

Parameters μ and N are related to the algorithm step size and the projection order respectively, and $\mathbf{x}(n)$ is a column vector composed by the last L samples of the input signal. Thus adaptive filters have L coefficients. Finally vector $\epsilon(n)$ is defined as

$$\epsilon(n) = \mathbf{A}^T(n)[\mathbf{A}(n)\mathbf{A}^T(n) + \delta\mathbf{I}]^{-1}\mathbf{e}(n). \qquad (3)$$

A simple adaptation rule that varies the projection order depending on the working state of the algorithm between a maximum and a minimum

values is needed. Equation (2) describes the variation of the filter weights, where the parameter $\epsilon(n)$ gives a measure of the size of this change. An estimation of the mean value of the weight changes can be used into the adaptation rule of N since it is closely related to the algorithm working state. The evolution of the projection order in [9] is directed by the instantaneous squared value of $e(n)$, which is compared with two given thresholds. However these thresholds are calculated from knowledge of the noise variance that has to be estimated either a priori or on line, which limits algorithm autonomy. An estimation of the mean squared value of $\epsilon(n)$ is also proposed to control variations in convergence step size in [2]. It is shown that this variation rule guarantees that the mean square deviation of the filter weights undergoes the largest decrease between algorithm iterations. Thus, the variation rule for the algorithm step size is proposed to be given by [2]:

$$\mu(n) = \mu_{max} \frac{\|\mathbf{p}(n)\|^2}{\|\mathbf{p}(n)\|^2 + C}, \tag{4}$$

where $p(n)$ is an estimation of the mean value of $\epsilon(n)$, which is obtained from an exponential weighting of its instantaneous value as $\mathbf{p}(n) = \alpha\mathbf{p}(n-1) + (1-\alpha)\epsilon(n)$ (with $0 < \alpha < 1$), and C is a positive parameter that depends on the algorithm projection order. It should be noted that this parameter is shown to be approximated by $\frac{N}{\text{SNR}}$ in [2], thus it has to be adjusted following the projection order values. The maximum step-size parameter μ_{max} in (4) is chosen to guarantee both fast convergence speed and filter stability and ideally should be less than 1 [4][12].

It seems clear from (4) that the step size will keep close to its maximum value when changes of the filter weights are high enough, which means that the algorithm has not achieved its steady state. On the other hand, small

values of weight changes will lead to a decrease in step size. The affine projection order could also be changed bounding the step size, thus, maximum and minimum threshold values using the step size can be defined as $\mu_{\max}\mu_{N\text{up}}$ and $\mu_{\max}\mu_{N\text{down}}$, with $0 < \mu_{N\text{down}} < \mu_{N\text{up}} \leq 1$. Consequently the simple rule of variation of the projection order can be given by

$$
N(n+1) = \begin{cases}
\min\left\{N(n)+1, N_{\max}\right\}, & \mu(n) > \mu_{\max}\mu_{N\text{up}} \\
N(n), & \text{other} \\
\max\left\{N(n)-1, 1\right\}, & \mu(n) < \mu_{\max}\mu_{N\text{down}}
\end{cases}
\tag{5}
$$

This way, the proposed algorithm would not only adapt the convergence step size $\mu(n)$ trying to minimize the residual error or accelerate convergence, but also would adapt the projection order to adjust the convergence and computational needs of the algorithm to its convergence state and therefore enhance the effect of the variable convergence step size.

3. Steady-state convergence analysis

The objective of this section is to evaluate the steady-state mean-square error performance of the proposed APA algorithm. The steady-state mean-square error (MSE) is defined as

$$
\text{MSE} = \lim_{n \to \infty} \text{E}\{|e(n)|^2\},
\tag{6}
$$

where $\text{E}\{\cdot\}$ denotes mathematical expectation and $e(n) = d(n) - \mathbf{x}(n)^T\mathbf{w}(n-1)$ is the output error at time n (note that $e(n)$ is the top entry of error vector $\mathbf{e}(n)$ in (1)).

It has been shown in [6] that the APA algorithm exhibits the following steady-state MSE

$$\mathrm{MSE_{APA}} = \lim_{n\to\infty} \left(\frac{\mu\sigma_r^2}{2-\mu}\right)\left(\frac{\mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\})}{\mathrm{Tr}(\mathbf{D}\cdot\mathrm{E}\{\boldsymbol{\Phi}(n)\})}\right) + \sigma_r^2 \qquad (7)$$

where matrices $\boldsymbol{\Phi}(n)$ and $\mathbf{D}$ are given, respectively, in (A.3) and (A.4) ($\mathrm{Tr}(\cdot)$ denotes the trace of a matrix).

The minimum MSE in (7) is reached when $N = 1$ as,

$$\mathrm{MSE_{APA=1}} = \frac{\mu\sigma_r^2}{2-\mu} + \sigma_r^2 \le \mathrm{MSE_{APA}}. \qquad (8)$$

It can be assumed that the proposed APA algorithm with the suitable threshold values reaches the steady state when $N = 1$. A choice of $\mu_{N\mathrm{down}}$ that guarantees the projection order will undergo a decrease up to 1 at steady state is given by (see motivations in Appendix B)

$$\mu_{N\mathrm{down}} > \frac{2}{2 + C\cdot\mathrm{SNR}(2-\mu_{\max})}. \qquad (9)$$

Furthermore the proposed APA algorithm dynamically changes the step-size parameter. Then, from (8), the MSE can be estimated as

$$\mathrm{MSE}_{proposed\ APA} = \frac{\overline{\mu}_\infty\sigma_r^2}{2-\overline{\mu}_\infty} + \sigma_r^2, \qquad (10)$$

where $\overline{\mu}_\infty$ is calculated in Appendix A as

$$\overline{\mu}_\infty = \lim_{n\to\infty}\mathrm{E}\{\mu(n)\} = \frac{(1+k) - \sqrt{(1+k)^2 - 2\mu_{\max}}}{k}, \qquad (11)$$

and

$$k = \frac{C}{N\sigma_r^2\mathrm{Tr}(\boldsymbol{\Phi}(n))} \approx \frac{C\cdot\mathrm{SNR}}{N}. \qquad (12)$$

Since $\overline{\mu}_\infty \le \mu_{\max}$, the proposed APA reduces the MSE of the APA algorithm with $\mu = \mu_{\max}$. Moreover, only the solutions that subtract the square root term in (11) should be considered. Figure 1 shows the evolution of $\overline{\mu}_\infty$ depending on k and $\mu_{\max}$. By substituting these values in (10) it leads to the corresponding theoretical MSE of the proposed APA algorithm.

4. Results

The proper performance of the proposed algorithm has been contrasted through various simulations, comparing it with the APA of fixed projection order and step size, and with the APA of variable step size proposed in [2]. Although the goal of the proposed algorithm is to improve both the computational performance and the residual level of the error signal in steady state, we must ensure that the transient state remains just as efficient as the other algorithms. Therefore, invariant and variant environments have been simulated. Thus the ability of the algorithms to adapt to changes in working conditions can be assessed.

An invariant system identification problem was chosen as first experiment. A plant modeled by a FIR filter of 20 coefficients (random) was used as the unknown system. The length of the adaptive filters was fixed to 19 coefficients, therefore a residual background noise was always present since the adaptive filters cannot achieve the exact model of the plant. The projection order of the APA was $N = 10$, which was also the maximum and initial projection order of the APA with adaptive projection order. The value of $\mu_{\max}$ was set at 0.08 and the thresholds values were $\mu_{N\text{up}} = 1/3$ and $\mu_{N\text{down}} = 1/2$. A typical value for α (α=0.99 as in [2]) in the estimation of mean value of $\epsilon(n)$ was used, and C parameter in (4) was fixed to $N/1000$ (this is a SNR close to 30dB although other close values are also possible without remarkable changes in algorithm behavior). Under these conditions and for each algorithm 3000 simulations were performed and averaged, using as reference signal, $x(n)$, random white noise of zero mean. The learning curves given by $10\log\left[\dfrac{e^2(n)}{d^2(n)}\right]$ were calculated and are shown in Fig. 2. The

evolution of the computational cost of each algorithm and the evolution of the projection order of the proposed algorithm are also represented in Fig. 2. It can be appreciated that as the algorithm approaches the steady state the projection order decreases and, consequently, so does its computational cost. Moreover, the final residual error not only is not worsened by this strategy but is slightly improved over the variable step-size APA and significantly improved compared to the APA of fixed parameters N and μ. Furthermore, the last two adaptive strategies do not adapt their computational cost to the algorithm state.

Although the first experiment demonstrates the efficiency of the algorithm when the proposed adaptive system operates within an stationary environment and the variation of the projection order is monotone decreasing, it would be interesting to see if it can also adjust when faster convergence is needed and hence also increments of the projection order. Thus the previous experiment was repeated but the coefficients of the plant were randomly changed after 10000 iterations and again after 20000 iterations, so the adaptive system had to readjust its working parameters from its input signals. The setup of this experiment was the same as the previous one. In this second case it is intended to illustrate the ability of proposed algorithm to track changes in system conditions. Fig. 3 shows the learning curves, the evolution of the projection order of the proposed APA and the computational cost of the algorithms for this second experiment. As can be seen in Fig. 3, the proposed APA presents better performance in terms of computational cost and final residual error. However, this algorithm shows a somewhat slower convergence behavior during transients because its projection order cannot

9

achieve the value of the other algorithms, so it cannot reach their convergence speeds. This drawback could be improved simply by either better adjusting the threshold levels of the projection order changes or its step size, for instance automatically switching to the highest projection order when a fast transient is detected. Thus it would be guaranteed that the proposed algorithm would show a transient behavior as good as the other, although the number of operations in the transients would consequently be higher. Fig. 4 presents the results of the experiment but with the above conditions. The thresholds have been changed to $\mu_{N\mathrm{up}} = 1/2$ and $\mu_{N\mathrm{down}} = 1/4$ and when N has to grow, it automatically switches to its maximum value ($N_{\max} = 10$ in this case). As can be appreciated from Fig. 4, the proposed algorithm improves its behavior within transient periods but it does not performs as well as before at steady state. Furthermore, the number of operations does not decrease as much as in the case shown in Fig. 3. The residual error and the computational cost of the proposed algorithm are, however, still better than for the other two adaptive AP alternatives even in this case.

To prove the behavior of the algorithm with colored noise, Gaussian noise filtered through an AR1 filter ($y(n) = \sqrt{1 - 0.9^2}x(n) + 0.9y(n-1)$) has been also used as reference signal, getting satisfactory results as it is shown in Fig. 5, and finally, the results when the input signal is exponentially distributed (non Gaussian) are shown in Fig. 6. The behavior of the algorithm is very similar for all of them.

5. Conclusion

In this paper an APA with variable step size and projection order has been presented. This algorithm fulfils two objectives: first to adjust computing needs to the actual demand of the convergence state (fast convergence speed is needed in transient state and hence higher projection orders and high computational load, whereas the reverse is true at steady state), and second, to further minimize the final residual error at steady state, since it decreases with the projection order. Although the second objective was already addressed by the algorithms of variable step size, the proposed algorithm overcomes the performance of these algorithms and improves the computational cost without any loss of convergence speed. Even though the total computational cost is greatly improved by the fact that the algorithm is not always working with the maximum or initial projection order. This algorithm can reduce its cost even further by using fast strategies developed for the APA family [13].

References

[1] K. Ozeki and T. Umeda, An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties, *Electron. and Commun. Japan*, vol. J67-A, pp. 126–132, Feb. 1984.

[2] H. Shin and A. Sayed, Variable step-size NLMS and affine projection algorithms, *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[3] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, Fast projection algorithm and its step size control, *in Proc. ICASSP95*, vol. 2, pp. 945–948, May 1995, Detroit, MI.

[4] C. Paleologu, J. Benesty, and S. Ciochina, A variable step-size affine projection algorithm designed for acoustic echo cancellation, *IEEE Trans. Audio, Speech and Language Process.*, vol. 16, no. 8, pp. 1466–1478, Nov. 2008.

[5] L. Liu, M. Fukumoto, S. Saiki, and S. Zhang, A variable step-size proportionate affine projection algorithm for identification of sparse impulse response, *Eurasip J. Advan. Signal Process.*, pp. 1–10, 2009.

[6] H. Shin and A. Sayed, Mean-square performance of a family of affine projection algorithms, *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 90–102, Jan. 2004.

[7] M. Ferrer, M. de Diego, A. Gonzalez, and G. Piñero, Convex combination of affine projection algorithms, *in Proc. 17th Eur. Signal Process. Conf.*, Aug. 2009.

[8] S. J. Kong, K. Y. Hwang, and W. J. Song, An affine projection algorithm with dynamic selection of input vectors, *IEEE Signal Process. Lett.*, vol. 14, no. 8, pp. 529-532, Aug. 2007.

[9] 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.

[10] F. Albu, C. Paleologu and J. Benesty, A Variable Step Size Evolutionary Affine Projection Algorithm, *in Proc. IEEE ICASSP2011* , pp. 429-432, May 2011, Prague, Czech Republic.

[11] S. Werner and P. S. R. Diniz, Set-membership affine projection algorithm, *IEEE Signal Process. lett.*, vol. 8, no. 8, pp. 231–235, Aug. 2001.

[12] S.G. Sankaran and A.A. Louis Beex, Convergence behavior of affine projection algorithms, *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 1086–1096, Apr. 2000.

[13] S.L. Gay and S. Tavathia, The fast affine projection algorithm, *in Proc. ICASSP95*, vol. 5, pp. 3023–3026, May 1995, Detroit, MI.

Biographies

Alberto Gonzalez was born in Valencia, Spain, in 1968. He received the Ingeniero de Telecomunicacion degree from the Universidad Politecnica de Catalonia, Spain in 1992, and Ph.D degree from the Universitat Politècnica de València (UPV), Spain in 1997. His dissertation was on adaptive filtering for active control applications. From January 1995, he visited the Institute of Sound and Vibration Research, University of Southampton, UK, where he was involved in research on digital signal processing for active control. He

is currently heading the Audio and Communications Signal Processing Research Group (www.gtac.upv.es) that belongs to the Institute of Telecommunications and Multimedia Applications (i-TEAM, www.iteam.es). Dr. Gonzalez serves as Professor in digital signal processing and communications at UPV where he heads the Communications Department (www.dcom.upv.es) since April 2004. He has published more than 70 papers in journals and conferences on signal processing and applied acoustics. His current research interests include fast adaptive filtering algorithms and multichannel signal processing for communications and 3D sound reproduction.

Miguel Ferrer was born in Almería, Spain, in 1974. He received the Ingeniero de Telecomunicacion degree from the Universitat Politècnica de València in 2000 and the Ph.D degree in 2008. In 2000, he spent six months at the Institute of aplicated research of automobile in Tarragona (Spain) where he was involved in research on Active noise control applied into interior noise cars and subjective evaluation by means of psychoacoustics study. In 2001 he began to work in GTAC (Grupo de Tratamiento de Audio y Comunicaciones) that belongs to the Institute of Telecommunications and Multimedia Applications. His area of interests includes efficient adaptive algorithm and digital audio processing.

Maria de Diego was born in Valencia, Spain, in 1970. She received the Ingeniero de Telecomunicacion degree from the Universitat Politècnica de València (UPV) in 1994, and the Ph.D degree in 2003. Her dissertation was on active noise conformation of enclosed acoustic fields. She is currently working as Associate Professor in digital signal processing and communications. Dr. de Diego has been involved in different research projects including

active noise control, fast adaptive filtering algorithms, sound quality evaluation, and 3-D sound reproduction, in the Institute of Telecommunications and Multimedia Applications (iTEAM) of Valencia. She has published more than 40 papers in journals and conferences about signal processing and applied acoustics. Her current research interest include multichannel signal processing and sound quality improvement.

Gema Piñero was born in Madrid, Spain, in 1965. She received the Ingeniero de Telecomunicacion degree from the Universidad Politecnica de Madrid in 1990, and the Ph.D. degree from the Universitat Politècnica de València in 1997, where she is currently working as Associate Professor in digital signal processing and communications. She has been involved in different research projects including array signal processing, active noise control, sound quality, psychoacoustics and wireless communications in the Audio and Communications Signal Processing (GTAC) group of the Institute of Telecommunications and Multimedia Applications (iTEAM) of Valencia. Since 1999 she has leaded several projects on sound quality evaluation for the automotive industry and currently her research interest includes sound quality applications in toys. Since 2001 she has been involved in several projects on 3G wireless communications supported by the Spanish Government and Telefonica. She has also published more than 40 contributions in journals and conferences about signal processing and applied acoustics. Her current research interests in the communications field include array signal processing for wireless communications, space-time coding and MIMO multi-user techniques. She is member of the IEEE since 1994 and founding member of the Spanish Association for Research and Teaching UNIVERSITAS.
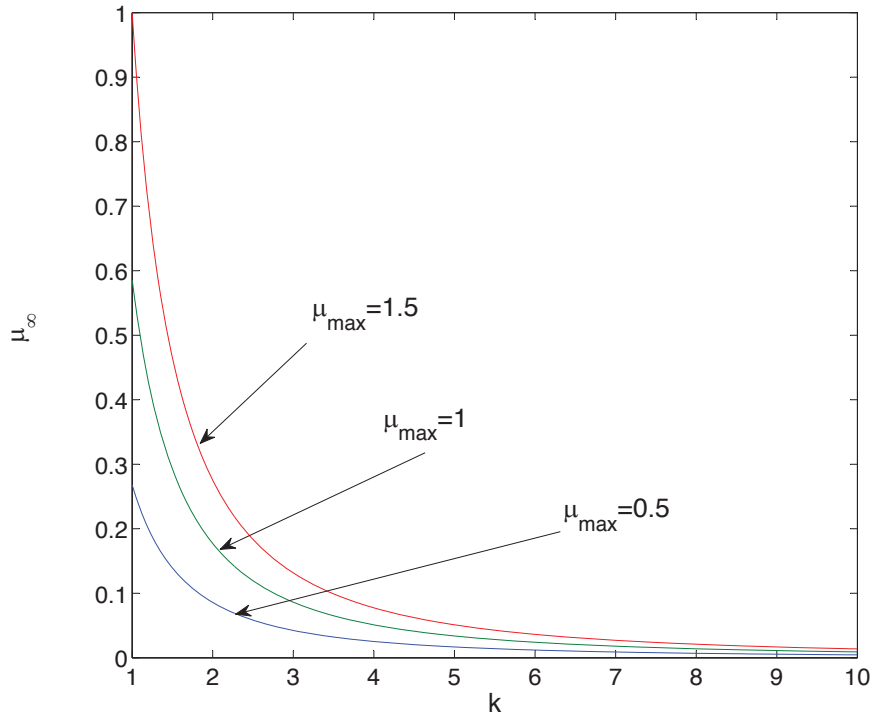
Figure 1: Estimated steady-state step size ($\mu_\infty(n)$) for different $\mu_{\max}$ values.

Appendix A. Computation of μ_∞

The data vector $\mathbf{d}(n)$ can be modeled as the output of an unknown system [6]

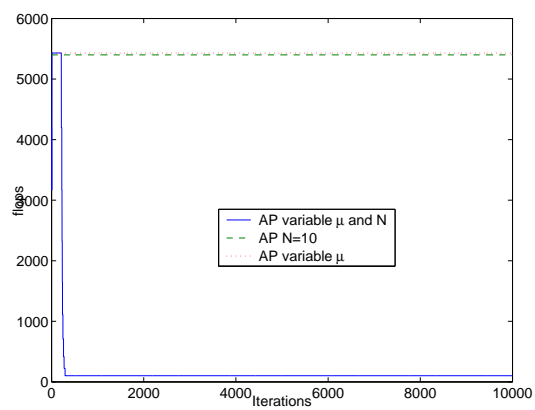$$\mathbf{d}(n) = \mathbf{A}(n)\mathbf{w}^o + \mathbf{r}(n), \tag{A.1}$$

where $\mathbf{w}^o$ is the vector we wish to estimate, $\mathbf{r}(n)$ is an N dimensional vector comprised of samples of the measurement noise r(n) with σ_r^2 variance and zero mean. The independence assumption between the different random variables that appear in the theoretical analysis is considered. Specifically, the variable
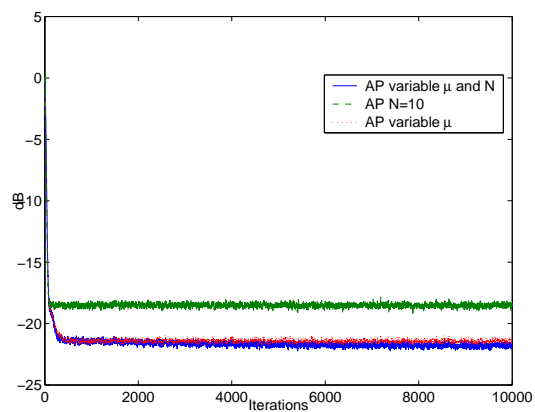
(a)
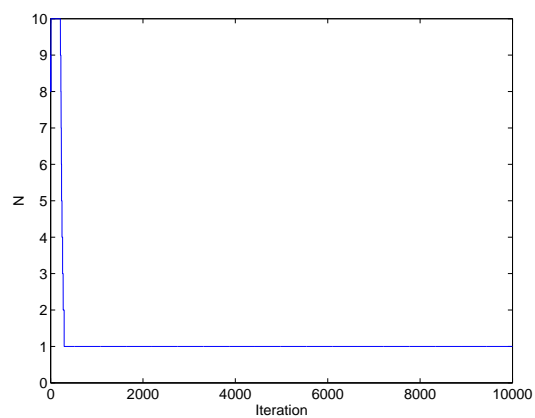
(b)

(c)

Figure 2: Learning curves (a), evolution of N (b) and computational cost (c) for an invariant system identification experiment (white noise).
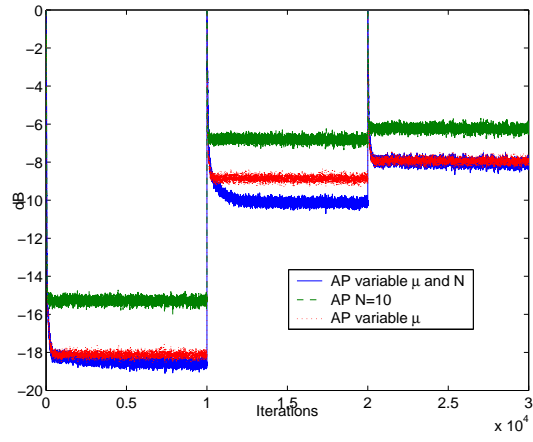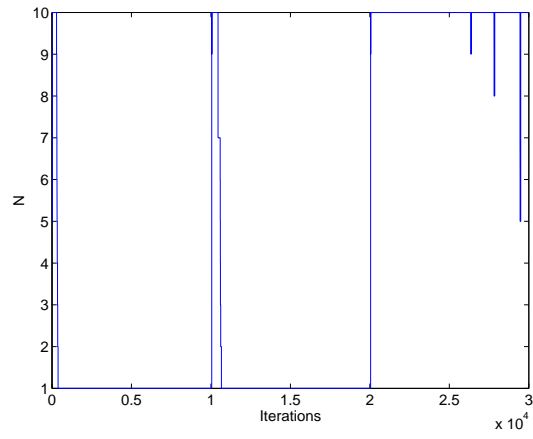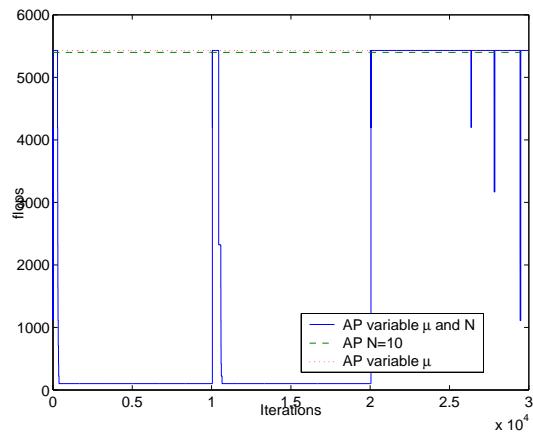
(a)

(b)

(c)

18
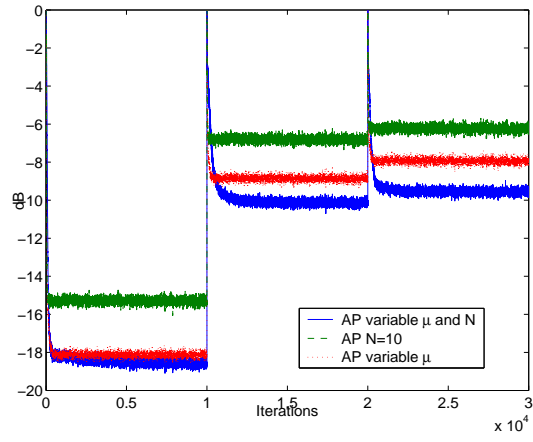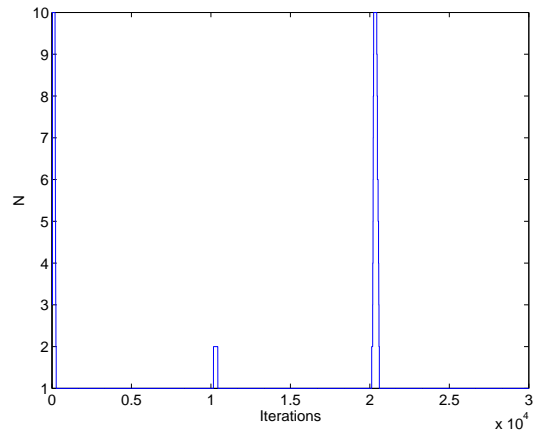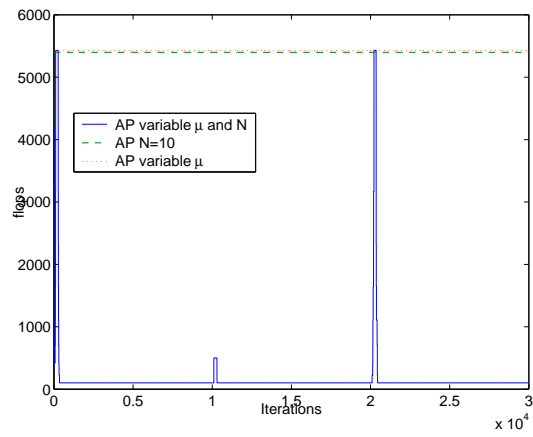
Figure 3: Learning curves (a), evolution of N (b) and computational cost (c) for a variant system identification experiment (original thresholds).
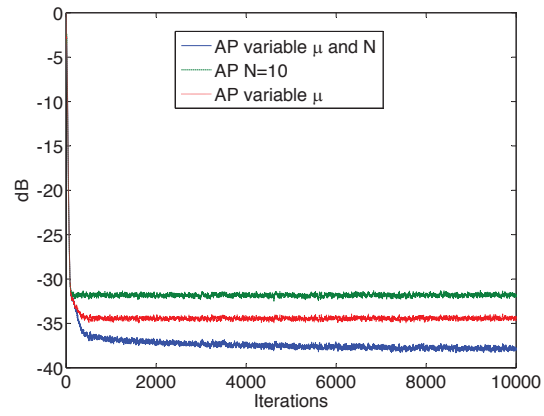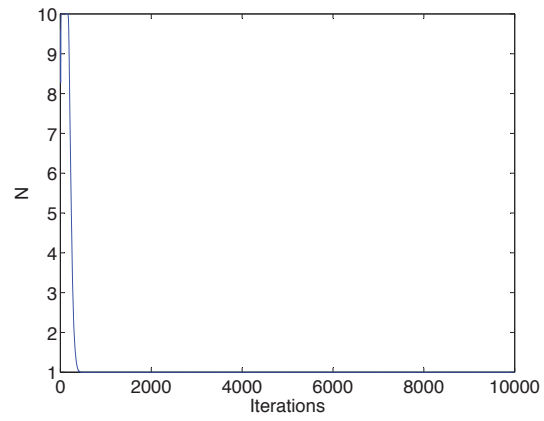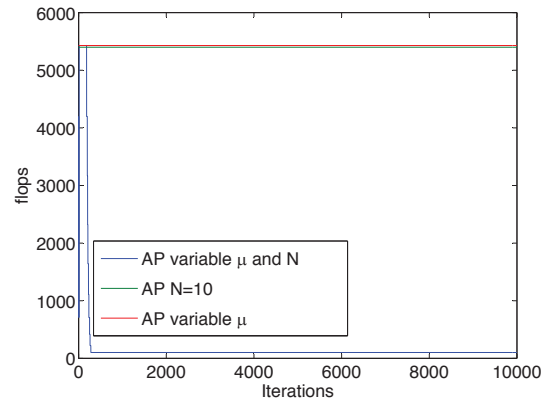
Figure 4: Learning curves (a), evolution of N (b) and computational cost (c) for a variant system identification experiment (modified thresholds).
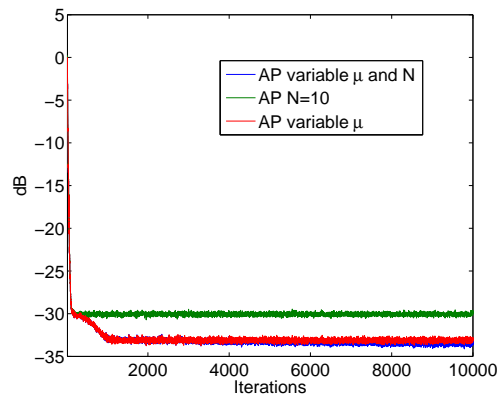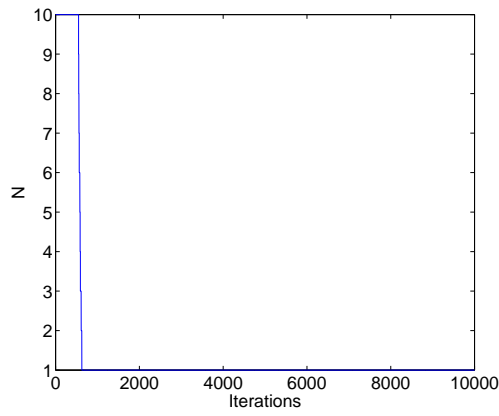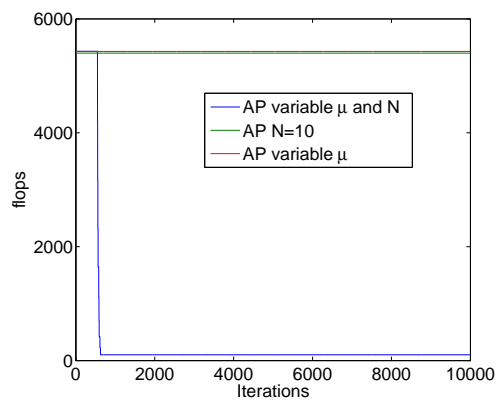
(a)

(b)

(c)

Figure 5: Learning curves (a), evolution of N (b) and computational cost (c) for a system identification experiment (colored noise).

20

(a)

(b)

(c)

21

Figure 6: Learning curves (a), evolution of N (b) and computational cost (c) for a system identification experiment (non Gaussian source signal).

step-size parameter $\mu(i)$ is considered to be statistically independent of the remaining random variables. Similarly to the steady-state MSE of the AP algorithm studied in [6], we find that the steady-state MSE of the proposed AP is given by ($\delta \approx 0$)

$$\text{MSE}_{\text{AP}} = \lim_{n \to \infty} \left(\frac{\mu(n)\sigma_r^2}{2 - \mu(n)} \right) \left(\frac{\text{Tr}(\text{E}\{\boldsymbol{\Phi}(\text{n})\})}{\text{Tr}(\text{E}\{\mathbf{D}(\mathbf{n})\boldsymbol{\Phi}(\text{n})\})} \right) + \sigma_r^2 \qquad (\text{A.2})$$

where the projection order at steady state is not limited to $N = 1$ and matrices $\boldsymbol{\Phi}(n)$ and $\mathbf{D}(n)$ are given, respectively, by

$$\boldsymbol{\Phi}(n) = [\mathbf{A}(n)\mathbf{A}^T(n)]^{-1} \qquad (\text{A.3})$$

and

$$\mathbf{D}(n) = \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ 0 & [1-\mu(n)]^2 & 0 & \cdots & \vdots \\ \vdots & 0 & [1-\mu(n)]^4 & & \\ & & & \ddots & 0 \\ 0 & & & & [1-\mu(n)]^{2(N-1)} \end{pmatrix}. \qquad (\text{A.4})$$

The first term on the right-hand side of (A.2) is known as the excess mean-square error (EMSE), which is defined by

$$\text{EMSE} = \lim_{n \to \infty} \text{E}\{|e^a(n)|^2\}, \qquad (\text{A.5})$$

where $e^a(n)$ is the top entry of $\mathbf{e}^a(n) = \mathbf{A}(n)[\mathbf{w}^o - \mathbf{w}(n-1)]$.

In order to evaluate $\overline{\mu}_\infty$, defined by

$$\overline{\mu}_\infty = \lim_{n \to \infty} \text{E}\{\mu(n)\}, \qquad (\text{A.6})$$

22

we consider the steady-state condition $p(n) = p(n-1) = \epsilon(n)$. Using $\mathrm{E}\|\mathbf{p}(n)\|^2$ instead of $\|\mathbf{p}(n)\|^2$ in (4), the steady-state step size becomes

$$\overline{\mu}_\infty = \lim_{n\to\infty} \mu_{max} \frac{\mathrm{E}\{\|\mathbf{p}(n)\|^2\}}{\mathrm{E}\{\|\mathbf{p}(n)\|^2\} + k \cdot N\sigma_r^2 \mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\})} \tag{A.7}$$

where the positive constant C is taken as a multiple of $N\sigma_r^2 \mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\})$ [2].

If we express $\mathrm{E}\{\|\mathbf{p}(n)\|^2\}$ in terms of $e^a(n)$, it holds the following approximation (see, e.g., [6] for details)

$$\begin{aligned}
\mathrm{E}\{\|\mathbf{p}(n)\|^2\} &= \mathrm{E}\{(\mathbf{e}^a(n))^T\boldsymbol{\Phi}(n)\mathbf{e}^a(n)\} + \mathrm{E}\{\mathbf{r}^T(n)\boldsymbol{\Phi}(n)\mathbf{r}(n)\} \\
&= \mathrm{Tr}(\mathrm{E}\{\mathbf{e}^a(n)(\mathbf{e}^a(n))^T\boldsymbol{\Phi}(n)\}) + \mathrm{Tr}(\mathrm{E}\{\mathbf{r}(n)\mathbf{r}(n)^T\boldsymbol{\Phi}(n)\}) \\
&\approx \mathrm{E}\{|e^a(n)|^2\}\mathrm{Tr}(\mathrm{E}\{\mathbf{D}(n)\boldsymbol{\Phi}(n)\}) + \mathrm{E}\{|r(n)|^2\}\mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\}).
\end{aligned} \tag{A.8}$$

Then, from (A.5), it yields

$$\lim_{n\to\infty} \mathrm{E}\{\|\mathbf{p}(n)\|^2\} = \mathrm{EMSE} \cdot \mathrm{Tr}(\mathrm{E}\{\mathbf{D}(n)\boldsymbol{\Phi}(n)\}) + \sigma_r^2 \mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\}). \tag{A.9}$$

By applying (A.9) in (A.7)

$$\overline{\mu}_\infty = \mu_{max} \frac{\mathrm{EMSE} \cdot \mathrm{Tr}(\mathrm{E}\{\mathbf{D}(n)\boldsymbol{\Phi}(n)\}) + \sigma_r^2 \mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\})}{\mathrm{EMSE} \cdot \mathrm{Tr}(\mathrm{E}\{\mathbf{D}(n)\boldsymbol{\Phi}(n)\}) + \sigma_r^2 \mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\}) + kN\sigma_r^2 \mathrm{Tr}(\mathrm{E}\{\boldsymbol{\Phi}(n)\})}. \tag{A.10}$$

If we consider $N = 1$, we obtain

$$\overline{\mu}_\infty = \mu_{max} \frac{\mathrm{EMSE} + \sigma_r^2}{\mathrm{EMSE} + \sigma_r^2 + k\sigma_r^2}. \tag{A.11}$$

where $\mathrm{E}\{\mathbf{D}(n)\boldsymbol{\Phi}(n)\} = \mathrm{E}\{\boldsymbol{\Phi}(n)\} = \mathrm{E}\{(\|\mathbf{x}(n)\|^2)^{-1}\}$. From (10), being $N = 1$, the EMSE is given by

$$\mathrm{EMSE} = \frac{\overline{\mu}_\infty \sigma_r^2}{2 - \overline{\mu}_\infty}. \tag{A.12}$$

By substituting (A.12) in (A.11), and solving the corresponding second order equation, it leads to the steady-state step size

$$\overline{\mu}_\infty = \frac{(1+k) \pm \sqrt{(1+k)^2 - 2\mu_{max}}}{k}. \tag{A.13}$$

Appendix B. Optimum $\mu_{N\text{down}}$

From (A.10), we can derive the optimum $\mu_{N\text{down}}$ in order to guarantee the algorithm decreases its projection order. Thus,

$$\frac{1}{\overline{\mu}_\infty} = \frac{1}{\mu_{\max}}\left[1 + \frac{kN\sigma_r^2\text{Tr}(\text{E}\{\mathbf{\Phi}(n)\})}{\text{EMSE}\cdot\text{Tr}(\text{E}\{\mathbf{D}(n)\mathbf{\Phi}(n)\}) + \sigma_r^2\text{Tr}(\text{E}\{\mathbf{\Phi}(n)\})}\right]$$

$$\text{(B.1)}$$

$$= \frac{1}{\mu_{\max}}\left[1 + \frac{kN}{\frac{\overline{\mu}_\infty}{2-\overline{\mu}_\infty} + 1}\right] \geq \frac{2 + kN(2 - \mu_{\max})}{2\mu_{\max}},$$

this leads to

$$\overline{\mu}_\infty \leq \frac{2\mu_{\max}}{2 + kN(2 - \mu_{\max})} \approx \frac{2\mu_{\max}}{2 + C\cdot\text{SNR}(2 - \mu_{\max})}, \qquad \text{(B.2)}$$

and finally

$$\mu_{N\text{down}} > \frac{2}{2 + C\cdot\text{SNR}(2 - \mu_{\max})}. \qquad \text{(B.3)}$$

This value guarantees the projection order decreases up to reach the steady-state with $N = 1$ and the corresponding estimated MSE is given in (10).