

A POLYPHASE IIR ADAPTIVE FILTER: ERROR SURFACE ANALYSIS AND APPLICATION

Phillip M. S. Burt

Max Gerken

LCS / DEE / EPUSP

Escola Politécnica - University of São Paulo, PO Box: 61548 São Paulo 01065-970 - Brazil
philip@lcs.poli.usp.br / mgk@lcs.poli.usp.br, Fax: (011) 818 5718 / Tel.: (011) 818 5129

ABSTRACT

An analysis of the local convergence speed of constant gain algorithms for direct form IIR adaptive filters is initially presented, showing the adverse effects that result from the proximity of the poles of the modelled system to the unit circle and, for complex poles, to the real axis. A global analysis of the reduced error surface in these cases is also presented, which shows that, away from the global minimum, there will be regions with an almost constant error, where the convergence of constant gain algorithms tends to be slow. A polyphase IIR adaptive filter is then proposed and its local and global convergence properties are investigated, showing it to be specially well suited for applications with underdamped low-frequency poles. The polyphase structure is tested with different constant gain algorithms in an echo-cancellation example, attaining a gain of 14 to 70 times in global convergence speed over the direct form, at the price of a relatively modest increase in computational complexity. A theorem concerning the existence of stationary points for the polyphase structure is also presented.

1. INTRODUCTION

IIR adaptive filters have the potential for providing considerably more efficient solutions than FIR adaptive filters in applications where the modelling of a system with highly underdamped poles (poles close to the unit circle) is required. This potential, however, hasn't yet been fulfilled. One of the major reasons for this is that simple constant-gain algorithms for IIR adaptive filters may converge very slowly to the optimum solution, for example, only in 10^4 to 10^6 iterations, or even worse. This can be observed for recursive gradient (RG) algorithms, Steiglitz-McBride Method (SMM) type of algorithms, SHARF and pseudo-linear regression (PLR) algorithms, alike. (The original references on these algorithms can be found in [1] and in [2]). Newton type algorithms, on the other hand, while having faster convergence, have a larger computational complexity than constant gain algorithms, which could take away much of the expected computational gain over FIR adaptive filters. This would be specially true if the input signal would allow a constant gain algorithm for the FIR case.

The purpose of this paper is twofold:

1) To present an analysis of the local convergence speed of constant gain algorithms for direct form IIR adaptive filters, showing the adverse effects that result from the proximity of the poles of the modelled system to the unit circle and, for complex poles, to the real axis. This analysis derives from the formulation in [1]. A global analysis of the reduced error surface in these cases is also presented,

which shows that, away from the global minimum, there will be regions with an almost constant error, where the convergence of constant gain algorithms tends to be slow. When considering the global convergence speed of such algorithms, this must be, therefore, taken into account, in addition to a low local convergence speed. This means that, unfortunately, in terms of convergence speed, the performance of simple direct form IIR adaptive filters is likely to be poor (using constant-gain algorithms) in the cases for which they attain a greater computational gain over FIR adaptive filters.

2) To propose a polyphase adaptive IIR filter which, in terms of the preceding analysis, exhibits better local and global properties than the direct structure. This structure can be adapted by any one of the algorithms used to adapt the direct structure, and in the case of systems with highly underdamped low-frequency poles, can attain, in exchange for a relatively modest increase in computational complexity, a much greater convergence speed than the direct structure.

2. LOCAL CONVERGENCE ANALYSIS

We consider that the system being modelled is $H(z) = C(z)/D(z)$, where $\deg[H(z)] = N$ is the minimum number of delays needed to implement $H(z)$, and that the adaptive filter has the form: $\hat{H}(z) = B(z)/A(z) = (b_0 + b_1z + \dots + b_Mz^M)/(1 + a_1z + \dots + a_Mz^M)$, $M = N$. Constant gain algorithms for IIR adaptive filters such as recursive gradient (RG), Steiglitz-McBride method (SMM) and pseudo-linear regression (PLR) have the general form

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{g}(n)e(n), \quad (1)$$

where vector $\mathbf{w}(n)$ contains the $2M + 1$ adaptive coefficients, $e(n) = [H(z) - \hat{H}(z)]u(n)$ is the output error for an input $u(n)$, μ is the adaptation gain and $\mathbf{g}(n)$ depends on the algorithm. At the global minimum $H(z) = \hat{H}(z)$, local convergence speed depends on the eigenvalue spread of the information matrix $\mathbf{I}(C, D) = E\{\mathbf{g}(n)\mathbf{g}^T(n)\}$ [1] which can be written as $\mathbf{I}(C, D) = \mathbf{S}(C, D)\mathbf{R}(D)\mathbf{S}^T(C, D)$, where $\mathbf{S}(C, D)$ is a resultant matrix formed by $C(z)$ and $D(z)$, and $\mathbf{R}(D) = E\{\mathbf{q}(n)\mathbf{q}^T(n)\}$, with $\mathbf{q}(n) = [u(n)/Q(z) \dots u(n-2M)/Q(z)]^T$. For RG and SMM algorithms $Q(z) = D^2(z)$ and for the PLR algorithm $Q(z) = D(z)$. The eigenvalue spread of \mathbf{I} , denoted by $\chi(\mathbf{I})$, is limited by $\chi(\mathbf{R})/\chi(\mathbf{SS}^T) \leq \chi(\mathbf{I}) \leq \chi(\mathbf{R})\chi(\mathbf{SS}^T)$ [1], and $\chi(\mathbf{SS}^T) < \infty$ if $C(z)$ and $D(z)$ are coprime. Therefore, if the roots of $C(z)$ are sufficiently far away from the roots of $D(z)$, the variation of $\chi(\mathbf{R})$ with $D(z)$ can serve as an

indicator of the variation of $\chi(\mathbf{I})$ with $D(z)$. Based on this, further analysis is aimed at the factors affecting the eigenvalue spread of \mathbf{R} .

Defining

$$\rho(i) = \frac{1}{\sigma^2} E \left\{ \frac{u(n)}{Q(z)} \frac{u(n-i)}{Q(z)} \right\}, \quad \sigma^2 = \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{1}{Q(e^{j\omega})} \right|^2 d\omega, \quad (2)$$

\mathbf{R} is a symmetric Toeplitz matrix whose first line is $\mathbf{r}_1 = \sigma^2 [1 \ \rho(1) \ \dots \ \rho(2M)]$. We consider the case when $D(z)$ has, among other roots, a pair of complex roots at $re^{\pm j\theta}$. It can be shown [6] that, as long as the spectral density of the input $u(n)$ has no zeros at $e^{\pm j\theta}$, as the roots approach the unit circle, the first line of \mathbf{R} , as would be expected from an intuitive reasoning, tends to

$$\lim_{r \rightarrow 1} \mathbf{r}_1 = \sigma^2 [1 \ \cos \theta \ \dots \ \cos 2M\theta], \quad (3)$$

which implies that the rank of $\lim_{r \rightarrow 1} \mathbf{R}$ is not greater than 2 [3]. When $H(z)$ has at least pair of complex poles, $M \geq 2$, and it follows that, as these poles approach the unit circle, the eigenvalue spread of \mathbf{R} tends to infinity, and local convergence speed tends to zero.

We now consider the effect of the angle θ on the eigenvalue spread of \mathbf{R} . Indicating the $2M$ eigenvalues of \mathbf{R} by $\{\lambda_i(\mathbf{R})\}$, and, for greater simplicity, considering $u(n)$ as unit power white noise, we have [3]

$$\min \{\lambda_i(\mathbf{R})\} = \min_{\|\mathbf{e}\|_2=1} \mathbf{e}^T \mathbf{R} \mathbf{e} = \min_{\|E(z)\|_2=1} \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{E(e^{j\omega})}{Q(e^{j\omega})} \right|^2 d\omega, \quad (4)$$

where $\mathbf{e} = [e_0 \ e_1 \ \dots \ e_{2M}]^T$ and $E(z) = e_0 + e_1 z + \dots + e_{2M} z^{2M}$. Now, since $\sum_{i=1}^{2M} \lambda_i(\mathbf{R}) = \text{tr}[\mathbf{R}] = 2M\sigma^2$ it follows that $\sigma^2 \leq \max \{\lambda_i(\mathbf{R})\} \leq 2M\sigma^2$, and

$$\chi(\mathbf{R}) \geq \left[\min_{\|E(z)\|_2=1} \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{E(e^{j\omega})}{Q'(e^{j\omega})} \right|^2 d\omega \right]^{-1}, \quad (5)$$

where $Q'(e^{j\omega}) = \sigma Q(e^{j\omega})$ and therefore $\|1/Q'(z)\|_2 = 1$. If an $E(z)$ can be found such that the integral above is low, it follows that the eigenvalue spread of \mathbf{R} must be large, due to the existence of the lower bound above. The effect of the angle θ can then be assessed by a frequency domain analysis of the requirements of $E(z)$ to fulfill such a condition: for $r \simeq 1$ and small values of θ , $|1/Q'(e^{j\omega})|$ is concentrated close to the origin, so that if all the roots of $E(z)$ are placed in the same region, $|E(e^{j\omega})/Q'(e^{j\omega})|$ will be low for any ω . For larger values of θ , $|1/Q'(e^{j\omega})|$ is not so concentrated close to the origin, and the roots of $E(z)$ have to be divided accordingly, resulting in values of $|E(e^{j\omega})/Q'(e^{j\omega})|$ which are not as low as previously. Therefore, it is to be expected that, for a given $r \simeq 1$, the eigenvalue spread of \mathbf{R} grows as θ tends to zero, that is, as the pair $re^{\pm j\theta}$ tends to the real axis. This can effectively be observed in practice.

3. ANALYSIS OF THE REDUCED ERROR SURFACE

In the following, input $u(n)$ is assumed to be white, though a similar analysis could be carried out for general inputs.

When the zeros of $\hat{H}(z)$ are such that the mean square error $\|H(z) - \hat{H}(z)\|_2^2$ is minimized for a given set of poles, we have [2]:

$$H(z) - \hat{H}(z) = g(z)V(z), \quad \text{where } g(z) = [V(z^{-1})H(z)]_{+}, \quad (6)$$

$[\cdot]_{+}$ being the strictly causal projection operator, and $V(z)$ being the all-pass function given by $V(z) = z^M A(z^{-1})/A(z)$. Note that $\|H(z) - \hat{H}(z)\|_2^2 = \|g(z)\|_2^2$, which, taken as a function of the parameters that establish the poles of the adaptive filter, is termed the *reduced error surface*.

Consider, for simplicity, that $\deg[C(z)] \leq \deg[D(z)] = N$, and that $H(z)$ doesn't have multiple poles. $H(z)$ can therefore be expanded in residues as $H(z) = K_0 + \sum_{i=1}^N r_i/(z - z_i)$. Using (6), we obtain, after some manipulations:

$$g(z) = \sum_{i=1}^N R_i \frac{z}{z - z_i}, \quad R_i = V(z_i^{-1}) \frac{r_i}{p_i}. \quad (7)$$

Using now this expression to calculate $\|g(z)\|_2^2$ by integration in z , we obtain:

$$\|g(z)\|_2^2 = \sum_{i=1}^N \frac{|R_i|^2}{|z_i|^2 - 1} + \sum_{j=1}^N \sum_{\substack{i=1 \\ R_i \neq R_j^*}}^N \frac{R_i R_j}{z_i z_j - 1}.$$

The case of poles close to the unit circle can now be introduced. Consider there are poles such that $|z_i| \simeq 1$ and $|R_i| \neq 0$, and that the expression above is dominated by m of these poles: $\|g(z)\|_2^2 \simeq \sum_{i=1}^m |R_i|^2 / (|z_i|^2 - 1)$. If the poles of $\hat{H}(z)$ are sufficiently far away from these poles of $H(z)$, then $|V(z_i^{-1})| \simeq 1$, since $|z_i| \simeq 1$ and $V(z)$ is all-pass. Therefore, under these conditions, a variation of the parameters that establish the poles of the adaptive filter has only a small effect on the mean square error $\|g(z)\|_2^2$, that is, we are in a flat region of the reduced error surface. As an example, we take the transfer function used in [4] to model an echo-path in HSDL loops, which, after normalization is $H(z) = (0.2178 - 0.8402z + 0.7971z^2)/(1 - 1.3148z + 0.3591z^2)$. The reduced error surface for this case is plotted in figure 1a), where the existence of a flat region away from the global minimum is clear.

Basically, the negative effect of this kind of error surface on the convergence of constant gain algorithms is that the steepness of the surface around the global minimum imposes a relatively low adaptation gain, which, in the flat region results in slow convergence. This effect is clear for gradient based algorithms. For other constant gain algorithms the effect is not as clear, though it can be argued that convergence is slower than in a steeper region, and this can effectively be observed in practice.

It should be noted that the existence of these regions is not only relevant when the initial values of the coefficients of the adaptive filter are far from their optimum values. For systems with poles close to the unit circle, even a relatively small error in the values of the coefficients can place them in a flat region of the reduced error surface.

4. A POLYPHASE IIR ADAPTIVE FILTER

The adaptive filter $\hat{H}(z)$ is now considered as having a polyphase structure:

$$\hat{H}(z) = \sum_{i=0}^{p-1} \hat{H}_i(z^p) = \frac{b_{P,0} + b_{P,1}z + \dots + b_{P,pM}z^{pM}}{1 + a_{P,1}z^p + \dots + a_{P,M}z^{pM}} \quad (8)$$

(where p is the polyphase expansion factor), which is indicated by writing $\hat{H}(z) = B_P(z)/A_P(z^p)$.

Initially, the question of eigenvalue spread is addressed. At the global minimum, $\mathbf{g}(n)$ in 1 is now a $(p+1)M+1$ vector and the information matrix can be written as

$$\mathbf{I}_P(C_P, D_P) = \mathbf{S}_P(C_P, D_P) \mathbf{R}_P^{(p)}(D_P) \mathbf{S}_P^T(C_P, D_P), \quad (9)$$

where $C_P(z)$ and $D_P(z^p)$ come from the polyphase form of $H(z)$. Matrix $\mathbf{S}_P(C_P, D_P)$ has $(p+1)M+1$ rows and $2pM+1$ columns and is not strictly the resultant matrix formed by $C_P(z)$ and $D_P(z^p)$, which we denote by $\mathbf{S}(C_P, D_P)$ and which would have $2pM+1$ rows instead and rank equal to $2pM+1 - \deg[T(z)]$, where $T(z)$ is the maximum common divisor of $C_P(z)$ and $D_P(z^p)$ [5]. If $C(z)$ and $D(z)$ have no common divisor then, $\deg[T(z)] = (p-1)M$ and the rank of $\mathbf{S}(C_P, D_P)$ would be $(p+1)M+1$. Though this remains to be proved we conjecture, based on several numerical examples, that after $(p-1)M$ rows of $\mathbf{S}(C_P, D_P)$ have been eliminated to obtain $\mathbf{S}_P(C_P, D_P)$ its rank is still $(p+1)M+1$. If this is always true, then $\mathbf{S}_P \mathbf{S}_P^T$ is non-singular when $C(z)$ and $D(z)$ have no common divisor, which means that $\mathbf{S}_P(C_P, D_P)$ would provide a suitable generalization of $\mathbf{S}(C, D)$ to the polyphase case.

As in section 2, we turn now to the analysis of $\mathbf{R}_P^{(p)}$. When $u(n)$ is white its elements are given by

$$\rho_P^{(p)}(i) = \frac{1}{\sigma_P^2} E \left\{ \frac{u(n)}{Q_P(z^p)} \frac{u(n-i)}{Q_P(z^p)} \right\}, \quad (10)$$

where $\sigma_P^2 = \int_0^{2\pi} |1/Q_P(e^{j\omega p})|^2 d\omega/2\pi$, and as previously, $Q_P(z^p) = D_P^2(z^p)$ for the RG and SMM algorithms and $Q_P(z^p) = D_P(z^p)$ for the PLR algorithm. A few manipulations permit to obtain the following property [6]:

Property 1 The elements $\rho_P^{(p)}(i)$ of $\mathbf{R}_P^{(p)}$, given by (10), take on the values given by

$$\rho_P^{(p)}(i) = \begin{cases} \frac{1}{\sigma_P^2} \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{1}{Q_P(e^{j\omega})} \right|^2 e^{j\omega m} d\omega, & i = mp, \\ 0 & i \neq mp. \end{cases} \quad (11)$$

We now obtain a matrix, denoted \mathbf{R}_P , by eliminating the null elements of $\mathbf{R}_P^{(p)}$, so that \mathbf{R}_P is a symmetric Toeplitz matrix whose first line $\mathbf{r}_{P,1}$ is given by

$$\mathbf{r}_{P,1} = \sigma^2 [1 \quad \rho_P(1) \quad \dots \quad \rho_P(2M)], \quad (12)$$

where $\rho_P(m) = \rho_P^{(p)}(mp)$. The following property can be proved [6]:

Property 2 The eigenvalue spread of matrix \mathbf{R}_P defined by (12) is equal to the eigenvalue spread of matrix $\mathbf{R}_P^{(p)}$ in (9).

This property is very useful since it means that the effects of the position of the roots of $D(z)$ on the eigenvalue spread of $\mathbf{R}_P^{(p)}$ for the polyphase filter can be investigated entirely by means of the effects of $D_P(z)$ on the eigenvalue spread of \mathbf{R}_P , which is exactly the same kind of problem already addressed in section 2. Now, the roots $z_{P,i}$ of $D_P(z)$ are related to the roots z_i of $D(z)$ by $z_{P,i} = z_i^p$, which means that $z_{P,i}$ is not as close to the unit circle as z_i , and, for complex roots, depending on the angle of z_i and the polyphase expansion factor p , $z_{P,i}$ will not be as close to the real axis as z_i . As seen in section 2 these both factors tend to reduce the eigenvalue spread and therefore increase the local convergence speed around the global minimum. When the modelled system has low-frequency poles close to the unit circle, for a convenient range of values of p the roots of $D_P(z)$, besides not being as close to the unit circle will also not be as close to the real axis as the roots of $D(z)$. Therefore, in these cases, the proposed polyphase adaptive filter structure is specially well suited as an alternative to the direct form adaptive filter.

We now consider the global properties of the reduced error surface. It can be shown [6], extending from the non-polyphase case, that the reduced error surface $\|g_P(z)\|_2^2$ in the polyphase case is determined by

$$g_P(z) = [V_P(z^{-p})H(z)]_+, \quad (13)$$

where $V_P(z^p) = z^{pM} A_P(z^{-p})/A_P(z^p)$. Following the same approach as in Section 2, when $\|g_P(z)\|_2^2$ has m dominating poles we may write $\|g_P(z)\|_2^2 \simeq \sum_{i=1}^m |R_{P,i}|^2 / (|z_i|^2 - 1)$, with $R_{P,i} = V_P(z_i^{-p})r_i/p_i$. Since z_i^p will not be as close to the unit circle as z_i , $|V_P(z_i^{-p})|$ will not be as constant as $|V(z_i^{-1})|$. The reduced error surface for which the coefficients of $A_P(z)$ are now taken into account will not be as flat away from the global minimum as in the direct form case. This can be seen in figure 1b), for the same echo-cancellation application considered previously.

An important property of error surface of the polyphase structure is given by

Theorem 1 In the case of white input and $\deg[H(z)] \leq M$, $\hat{H}(z)$ as given by 8 is a stationary point of $\|H(z) - \hat{H}(z)\|_2^2$, if, and only if, $H(z) = \hat{H}(z)$.

The demonstration of this theorem is in [6] and [7] and will not be presented here due to lack of space. It should be noted, however, that this property does not follow directly from similar and already known properties of the non-polyphase form of $\hat{H}(z)$, and therefore effectively constitutes a new result.

5. SIMULATION RESULTS

The already mentioned constant gain algorithms and a Newton version of the RG algorithm (denoted RG/N) were applied to the proposed polyphase adaptive IIR filter structure and to the direct form adaptive IIR filter, for the aforementioned HSDL echo-cancelling problem. An ensemble of 100 realizations of a gaussian white noise input signal was utilized in the more rapidly converging cases, while otherwise only the first 10 realizations of the ensemble were utilized. For each case, the mean number of iterations n_c needed to reach and maintain an output error of less than -60 dB in relation to the desired output was measured for increasingly higher values of μ (which for the RG/N algorithm corresponds to 1 minus the forgetting factor). The

selected values of μ were those which resulted in the lowest n_c while still assuring convergence for all of the tested realizations of the input. These results are presented in the following table and plots for the RG algorithm are showed in figures 1c) and 1d). The trade-offs involved in the selection of a particular algorithm are not under discussion here, the important point being that, for each algorithm, the polyphase structure converged faster than the direct form.

Structure	Algorithm	μ	n_c
Direct	RG	3.5×10^{-4}	1.7×10^5
	SMM	4.0×10^{-4}	8.8×10^4
	PLR	6.0×10^{-3}	1.4×10^4
	RG/N	5.0×10^{-2}	5.5×10^2
Polyphase ($p = 4$)	RG	1.0×10^{-2}	2.4×10^3
	SMM	1.0×10^{-2}	2.2×10^3
	PLR	2.0×10^{-2}	9.9×10^2

As it can be seen in the table above, the gain in convergence for the constant gain algorithms speed ranges from a factor of 14 to a factor of 70, depending on the specific algorithm.

The computational complexity was calculated for: a direct form adapted with RG and RG/N algorithms; the proposed polyphase structure adapted with the RG algorithm and a transversal FIR filter adapted with the LMS algorithm. The total number of multiplications, additions and divisions are listed in the following table, calculated for the orders involved in the example which was presented ($M = 2, p = 4$) and, based on [4], assuming a 100 tap FIR filter.

Structure	Algorithm	\times	$+$	\div
Direct IIR	RG	13	12	0
Polyphase IIR	RG	25	24	0
Direct IIR	RG/N	92	52	1
Trans. FIR	LMS	203	202	0

It can be seen that the increase in the convergence speed attained by the polyphase structure comes, in relation to the direct form adapted with a Newton algorithm, at a low price in terms of computational complexity, which is still much less than that of the FIR solution. It should be noted also that this benefit over a Newton algorithm increases with the degree of the filter.

REFERENCES

- [1] H. Fan, "A Structural View of Asymptotic Convergence Speed of Adaptive IIR Filtering Algorithms: Part I-Infinite Precision Implementation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp 1493-1517, April 1993
- [2] P.A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, New York, 1995.
- [3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall
- [4] W. Y. Chen et al., "High Bit Rate Digital Subscriber Line Echo Cancellation", *IEEE Journal on Selected Area in Communications*, pp 848-860, August 1991
- [5] P. Lancaster and M. Tismenestsky, *The Theory of Matrices*, Academic Press, 1985
- [6] P. M. S. Burt, Doctoral Thesis
- [7] P. M. S. Burt and M. Gerken, "A Polyphase Adaptive IIR Filter ", to be submitted to the *IEEE Transactions on Acoustics, Speech and Signal Processing*.

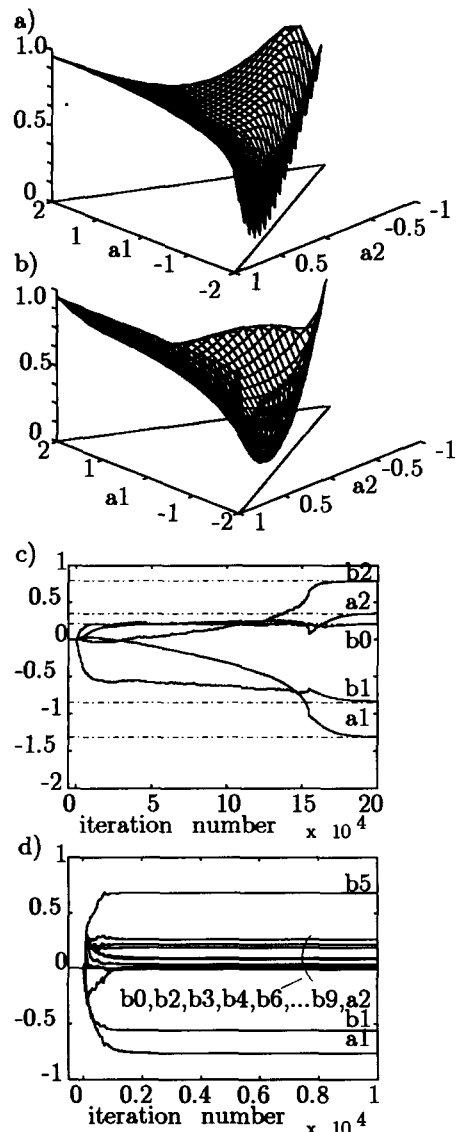


Figure 1. a), b) example of reduced error surface of direct form filter and polyphase filter, respectively; c),d) adaptation of coefficients of direct form filter and polyphase filter, respectively